UNCLASSIFIED

AD **281 847**

Reproduced
by the

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

UNCLASSIFIED

62-175

ON THE POSSIBLE AND IMPOSSIBLE IN OPTICS

By

G. G. Slyusarev

UNEDITED ROUGH DRAFT TRANSLATION

ON THE POSSIBLE AND IMPOSSIBLE IN OPTICS

BY: G. G. Slyusarev

English Pages: 163

S/5023

Date 16 July 1962

T

G. G. Slyusarev

O VOZMOZHNOM I NEVOZMOZHNOM V OPTIKE

Pages: 1-190

FTD-TT-62-175/1+2

TABLE OF CONTENTS

II

A strict solution of optical problems involving the propagation of light is both complex and cumbersome, and can only be found in certain cases. The main reason for this is the undular, electromagnetic nature of light.

We almost always have to limit ourselves to approximate methods and results. Geometric optics, which originated a long time ago, is a very crude and simplified method of solving optical problems. Study of the interference of light somewhat improved knowledge of the optics of light, although the methods were still approximate. The theory of the diffraction of electromagnetic waves has finally shown the path leading to an exact solution of optical problems. This path is so complicated and thorny, however, that the theory obviously has to be modified in the case of a number of simplified and approximate methods with varying degrees of accuracy. Furthermore, for the sake of simplicity optical problems are usually soved for the case of non-existent and impossible "points of light" radiated uniformly in all directions.

Given this state of affairs, it is not surprising that in optics, more often than in any other branch of physics, all kinds of paradox3s have to be dealt with at every step. It is not surprising that people who have mastered the approximate laws of geometric optics and are used to working with "luminous points" and "parallel beams" embark upon the invention of devices which reduce radiation from a luminous body to a point, or combine telescopes and microscopes in the hope that they will see new details on the sun and moon, and are amazed at the unimaginativness of experienced optical specialists who have not even thought of using one microscope to look at an image through another microscope in order to be able to see molecules!

The scientific establishments which engage in the study of optics are usually flooded with suggestions of this kind from all corners of the globe. Inventions involving perpetuum mobile are far less common nowadays than designs for ultra-ultramicroscopes, supermicroscopes and optical systems which can set fire to objects at a distance. The decrease in the number of suggestions for perpetual motion has undoubtedly been aided by popular science literature. Unfortunately, in the field of optics there has not been so far any literature giving a fairly simple and coherent account of what is possible and impossible, and it is now high time the ideas and energy of inventors were directed into more hopeful channels.

The appearance of this little book by Professor G.G. Sluysarev, and outstanding specialist in mathematical optics and author of the " Theory of the Calculation of Optical Systems" , is very timely. Naturally, within the space of this short book

Professor Sluynarev has certainly not been able to explain all the paradoxes to be found in optics, and much has had to be described rather briefly and sketchily. There will no doubt be need for more books and booklets with fuller explanations of three optical paradoxes, but at the present stage we must welcome the "first swallow of spring" in a much needed field.

1944   S. Vavilov

AUTHOR'S PREFACE

The spheres of application of optics are broadening and multiplying with every year. Engineers, chemists, doctors, biologists and military experts are all having to learn the fundamental principles and laws of this branch of physics. But optics is an abstract discipline and not always one easy to master. Although it is relatively easy to acquire a practical knowledge of the basic properties of optical instruments, the theoretical fubdamentals are mych harder to assimilate. That is why it is possible to make errors, sometimes extremely suble ones difficult to detect. most often the possibilities of the optical instruments are overrated and give rise to unrealizable hopes, and later -- after futile attempts -- to disappointments.

The opinion is current among many optical specialists, sometimes leading experts in the field, that optics is an all-powerful weapon which can make beams of light do anything at all -- produce any degree of illumination , (which is just tolerable) or any degree of brightness (which is contrary to the second principle of thermodynamics), make any beam parallel, and even turnit into a needle-point, which is just one step away from the fantasy of science fiction.

The blame for this misunderstanding of optical systems lies to a large extent with high-school textbooks, In many of then geometric optios, photometry and diffraction, which are closely interlinked and determine the properties of all optical instruments, are presented as completely independent branches[1].

From the standpoint of general physics this distinction is unimportant, but it affects the theory of optical instrunents: when split up into little pieces this branch of science soon vanishes fron the memory.

No matter how this state of affairs came about, we must do something to prevent its harmful consequences. The author hopes that this book may be of some use in this respect.

---

[1] We have found a number of incorrect statements in one particular upper-grade textbook used in high schools. It states, for example, that the magnification of telescopes ranges fron 1000 to 10,000 (this latter figure is clearly exaggerated) and that microscopic enlargement reaches as nuch as 3000 (such magnification is useless and even harmful). The textbook does not explain why there is a limit in magnification. In the chapter on diffraction there is no reference to the ways in which it is manifested n optical instruments, while it is actually the diffraction which prevents high magnification.

The reader is not offered a systematic course in optics, not even additional chapters to supplement those missing in elementary courses. The author has selected, more or less on an arbitrary basis, several problems of applied optics, in particular those which inventors find alluring (this includes optical experts as well as specialists in other branches of engineering), for example, the problem of being able to set fire to objects at a distance and super-magnification in telescopes and microscopes. It also sets forth some views on more subtle problems, such as "optical amplifiers" with a large scope angle, prospects for the development of optical instruments, and so forth. Although the book has no definite plan, all the inventions considered have one feature in common -- they are not possible in practice, either in the absolute sense or at least for a long time to come.

The author certainly does not wish to sow seeds of optical pessimism in the minds of his readers by limiting the framework within which new applications of optical instruments are commonly sought. Hence the book describes several new ways of increasing the power of optical instruments, which have been recently developed but are still little known, wven to specialists. These include the interpretation of deffraction patterns which relace images at high degree of magnification, and different ways of improving contrast of the image and the use of electron instead of luminous fluxes.

After the publication of the first edition of this book, there appeared a number of curious devices invented by reputable authors; these devices seemed feasible at first, but were actually based on a misinterpretation of the most important principles of optics. The misinterpretation arose from the fact that in the inventions the pr8nciples were masked by extremely complex designs which confused both the inventors themselves as well as the pitiable reviewers. Some of the more vivid "inventions" in this category are presented in a simplified form, divested of their cumbersone and futile trimmings; in this form it is easier to bet to grips with the underlying error in them.

Optical parts are sometimes considered to posses more dangerous properties than they actually have; this leads to undue precaution in dealing with them, particularly in respect of their fire hazard. This question is considered in a separate chapter. There is also a description of the phase contrast method which is now enjoying well-deserved popularity.

Despite the small scope of the book, and the problem considered, the author was nevertheless forced to condult a number of members of the State Optical Institute and takes the opportunity of expressing his gratitude to those concerned.

In particular the author considers his duty to point out that the idea of the book belongs largely to Professor S.I. Vavilov, who rendered a great deal of valuable assistance. The author is alsoindebted to Professor Ya.Ye. Ellengorn and V. Ye. Kozlov for the illustrations in the chapter on microscopes.

The second edition of the book was reviewed by Academician G.S. Landsberg who made sp,e cprrectopns and useful comments. To the author's deep regret the death of G.S. Landsberg prevents the author thanking him personally.

The third édition is the same as the second except for the addition of the section "Recovery of Sight" when the retina is damaged" and a new, simpler account of the phase contrast effect, borrowed from Tsernike.

<div align="right">G. Slyusarev</div>

November, 1959.

CHAPTER I

COMBUSTION FROM AFAR

Sec. 1. The origin of the notion of combustion from afar

The legend that Archimedes used mirrors to set fire to the enemy ships anchored off Syracuse has stimulated the imagination of many generations of inventors. It is noteworthy that the legend only became widespread in the Middle Ages; it seems as though one of the persons responsible for this was an inventor of the time who wished to be able to use the authority of Archimedes to justify his own efforts in the same direction. Jean Battista della Porta, the inventor of the "camera obscura", has left us a description of a combustion machine of this kind, almost identical in principle with the hyperboloid of Garin the engineer in A. Tolstoy's well-known story.

It should be noted that even as long ago as the Middle Ages there were scientists who did not believe in the possibility of such machines. In his "dioptrics" Kepler refuted Porta's hypothesis by means of arguments which coincide in content, if not in form, with those which we shall adduce.

An extensive contribution to the popularization of combustion machines has been made by different writers for whom these death-rays burning up everything in their path are a veritable treasure. The Martians in the novel by H. G. Wells almost conquered the earth by using weapons which emitted heat rays; Wells was too shrewd, however, to describe the weapons the Martians used. Engineer Garin, who was created by one of our own Soviet writers, is undoubtedly a pioneer in the invention of long-range combustion machines, and there is no doubt that he has had a great effect on the minds of many inventors who have made use of his basic idea. The literary craftsmanship of the celebrated writer, A. Tolstoy, made the idea seem true to life, although it is radically impossible, as we shall see. Furthermore, engineer Garin had obviously forgotten his geometry, since he confused hyperboloids with paraboloids, and did not know how to find the focus of the latter.

It is curious that Jules Verne, who predicted the technical developments of the Twentieth Century in many fields, avoided the trap which ensnares many writers of science fiction. His prolific writings contain descriptions of many adventurous investions. Optical instruments, as befits them, are also given a place. But non of Jules Verne's hero scientists ever thought of using heat ray. His great common sense cautioned him against too unlikely inventions.

As a result of adventure stories and ancient legends, the number of inventions which are supposed to cause fires or explosions by means of heat rays at great distances is increasing from year to year. It would appear that the tremendous development of engineering in the last few centuries gives us every reason to believe that this long dreamed of aim can be achieved, and it is quite natural to wonder why the solution has not yet been found.

Just as during the time of Archimedes, for most contemporary inventors the energy carrier is a beam of light rays concentrated by some sort of optical device. Since both sources of light rays and optical instruments have been perfected to a great extent, it is natural to believe that the solution of the problem is also not far off.

Let us examine the reasons why thousands of designs already put forward and thousands of others which will see the light of day in due course are doomed to failure, despite the complexity and cleverness of the arrangement of parts in the optical device. We are obviously disregarding instruments which are in direct proximity to the object which can set fire to objects from afar is quite feasible theoretically. If we have a powerful light source and an extensive optical system, we can concentrate enough energy to ignite any inflammable substance, such as paper or sawdust, from any distance away.

The heating effect here depends both on the source of the rays and the optical system transmitting the energy. But given the present day sources of light, we require optical systems of fantastic dimensions, as we shall see later from the theoretical equations.

There is a simple experiment, known to any schoolboy, that explains both

the basic principle of all inventions relating to combustion from afar, as well as the complete futility of them (when considered more closely).

Let us take THE CASE OF a condensing lens with the greatest possible diameter and shortest focal distance, for example, an ordinary magnifying glass, on a bright sunny day. We will position it so that the axis faces the sun. At its focal point the lens forms a tiny image of the sun, and if the ratio of the diameter and focal distance is fairly large (from now on this ratio is called the relative aperature of the lens), we can easily set fire to pieces of paper, cloth, cork and so on.

If we take a very long focus lens, let us say a spectacle lens of half a diopter (positive), we produce an image of the sun in the form of a large circle 2 cm in diameter about 2 m from the lens, but the image is so weak that it can be looked at on a sheet of white paper without discomfort. Obviously, there is no combustion in this case.

The effect of the relative aperture of the lens can easily be explained.

Sec. 2. Fundamentals of Study of luminous energy

First of all, let us agree that we will use the entire apparatus and terminology of photometry, since that science has been fully developed, is convenient and applicable to our problem, except that we will have to broaden the sphere of its application for our purposes.

Photometry studies the properties of visible light rays, i.e., rays which cause a reaction in the retina of our eyes. But we shall be dealing not only with visible, but also invisible infra-red rays. To avoid introducing unnecessary new units, we will use the conventional photometric units and give them slightly different values corresponding to a wider region of the spectrum.

Despite the great possibility that the uninitiated readers will have to reach for their photometry textbooks*, we will outline here the fundamentals required for further argumentation, especially since for our purpose we will describe them from a slightly different angle.

A beam of rays carries a certain amount of energy, the presence of which can easily be detected by intercepting them, for example, with a soot-covered screen. The screen heats up under the action of the beam until a thermal balance is established,

i.e., until the beam emitted by the heated body becomes equal to the stream impinging upon it. The heatup depends, on the one hand, on the incident stream, and, on the other, on the properties OF THE ENERGY RECEIVER AND ITS POWER to absorb that energy. For example, a freshly applied layer of magnesium, plaster of paris or silver reflects 95 - 98% of the incident energy. A surface coated with soot or velvet absorbs 96 - 99% of the incident energy.

The ratio between the amount of light reflected in all directions by a mat surface and the amount of light which originally impinged upon it is termed the reflecting power or albedo. The albedo a varies from zero to unity.

What is the relationship between the heating up of a body which absorbs all light energy impinging on it, and the stream of energy itself? As long as the temperature of the body is still a long way from the equilibrium point, the heating up is proportional to the energy of the flux per unit area (i.e., the less the area over which the flux is concentrated and the longer the duration of the heating, the more the body heats up).

The sun radiates energy with a power of 2 cal/min per cm^2, i.e., an area of 1 cm^2, for which $a = 0$, illuminated by direct sunlight, absorbs 2 calories a minute, or 0.14 watts. If the heat capacity of the area is unity, the HEIGHT is 1 cm and the thermal conductivity is very high (infinite), the area heats up 2° per minute, or 120° per hour. Any black body heated to 120°C radiates 0.14 watts from 1 cm^2 of its surface. Hence light of the sun cannot raise the temperature of any body above 120°C.

Experiments show that a temperature of 500 - 700°C is required to ignite dry wood. This temperature could only be obtained by a black body if every square centimeter of surface received a flux of 2 - 5 watts, i.e., 20 - 40 times more than the sun provides under the best circumstances; the heating would have to be continued for some time as well - 20 or 30 minutes.

_____

[*] For example, easily understandable textbooks "General Introduction to Photometry" by Fabry, Gostekhizdat, 1934, or S. O. Meisel's "Light and Vision", Voenizdat, 1949.

An output of hundreds of watts is required for instantaneous ignition, since most easily inflammable solids absorb only 30 or 40% of the energy impinging upon them. A flux of 150 watts, 50 watts of which would be used to heat up the body, could raise the temperature of it to 1500°C. It would only need a few seconds to set fire to paper, planks of wood and so on.

More detailed information on the effect of radiation of different intensities is given in Table 1; it has been borrowed from a translation into Russian of Lawson's booklet "The Atomic Bomb and Conflagrations" (Foreign Lit. Press, 1955, p. 12).

We are obviously interested in ways of producing a flux of sufficient power to set fire to easily inflammable bodies long distances away.

Let us consider a source of light sending out rays over certain area of space. In this area of space let us look at the so-called light beam, i.e., an elementary volume filled with luminous energy.

Table 1.

Impression or effect	Intensity in cal/cm^2. sec.
Sunlight in summer......................................	0.016
Discomfort after 3 seconds..............................	0.25
Lower limit of radiation sufficient to ignite any kind of wood after prolonged action if a small flame is placed half an inch from the surface in addition to radiation from afar,..............................	0.3
Lower limit of radiation sufficient to cause spontaneous combustion in any type of wood after prolonged action.........................	0.7
Spontaneous combustion of cotton cloth...	
In 7 seconds...	0.8
In 5 seconds...	1.0
In 3 seconds...	1.3
Spontaneous combustion of insulated fiber-pasteboard	
In 7 seconds...	1.1
In 5 seconds...	1.25
In 3 seconds...	1.4
Spontaneous combustion of thick oak planks	
In 20 seconds..	1.1
In10 seconds..	1.3
In 8 seconds...	1.35

If the medium through which the light spreads is isotropic (has identical properties in all directions)- which is always the case in practice - the light beams are bounded by linear surfaces. A beam of this kind is bounded by two areas s and s'

(Fig. 1); in a particular case s may be part of the surface of the source and s' may be part of the surface of the receiver.

Let us first consider the general case in which s does not coincide with the source, nor s' with the receiver.

Let s be a small area radiating energy and let ON be the normal to this area. Let s' be another small area perpendicular to the direction OO'. Let us consider the total luminous flux radiated by the area s and received by the area s'. Since these areas are very small, we may take it that the beam is homogeneous within the cone of rays passing through s and s'. Hence the flux through these areas is proportional to s and s'. Furthermore, if the direction of the beam axis OO' forms the angle θ with the normal ON to the area, the flux is also proportional to $\cos \theta$, since it is only the projection of the area s' onto the plane perpendicular to OO' which acts in the direction OO'.
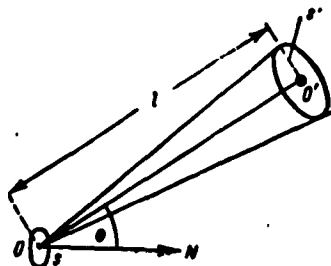
Fig. 1. Light beam.

Let us now consider the effect of the distance l between s and s'. If we move the area s' from s to a distance n times greater, it is clear that the power on it is n^2 times smaller. Hence the power is inversely proportional to l^2 and for the power $\bar{\Phi}$ we can write down the equation

$$\Phi = \frac{Bss' \cdot \cos \theta}{l^2},$$

where B is a coefficient depending on the properties of the source, but not depending on the shape or boundaries of the beam. The coefficient B is termed the brightness of the beam: this concept may be applied as well to the part of the source surface intersected by the beam, provided the latter is extended as far as the source. If the source is not a surface one, but a body source (e.g., a luminous gas, the sky, and so on), its brightness is determined on the basis of the equation given for the flux.

The equation for $\overline{\Phi}$ can be written in the form of the product of two

multiples

$$\Phi = (Bs \cdot \cos \theta)\left(\frac{s'}{l^2}\right).$$

The first multiple is conventionally termed the light intensity of the
area s in the direction s'; the second multiple is merely the solid angle, at which
the area s' is visible from the element s, or leaving aside for the moment the area
s' which was only used to illustrate the point, the solid angle of radiation. Thus,
the elementary flux $\overline{\Phi}$ is the product of the light intensity I in the direction of
radiation and the solid angle of radiation ω

$$\Phi = I\omega.$$

If the source is large in size and the solid angle of radiation ω is large,
the total flux is the sum of all the elementary fluxes obtained by dividing the area
and solid angle into elements.

Attention should be given to the particular case in which s is part of the
surface of the luminous source. Any luminous body emits light from different points
and in different directions. Generally speaking, its brightness varies from point to
point and also varies at each point according to the direction of radiation; here the
relationship between brightness and direction can be represented by any law.

The equation given for $\overline{\Phi}$ remains valid for this case, but B has to be
taken in accordance with the law of radiation, while $\cos \theta$ is considered equal to
unity. The flux is expressed in lumens, the light intensity in candles, and the
brightness in nits if we take the meter as the unit of length, and in stilbs if we
take the centimeter as the unit of length.

The effect of a beam of light on a screen placed in its path is determined
by the density of the flux, which photometrists call the illumination, and this is
equal to $E = \dfrac{\Phi}{s'}$, where s' is the area of the illuminated element.

The rise in temperature over a fairly brief period of time is also
proportional to the absorptivity of the screen. E is expressed in phots. The phot

is often replaced by the lux (1 lux = 10^{-4} phots).

As an illustration, let us look at the brightness (in the visible region of the spectrum) of some light sources. The brightness of the sun is 120,000 stilbs, an electric arc crater is 15,000 stilbs, a "point" bulb ranges up to 2,500 stilbs, and the filament of an incandescent bulb has a brightness of 300 - 800 stilbs.

The illumination (in the visible region of the spectrum) due to the sun on the earth's surface is about 10 phots.

We can now determine the illumination required to ignite inflammable material, for example in the experiment with the magnifying glass described above.

Let us consider two areas: one is directly illuminated by sunlight while the other is illuminated by the image of the sun produced by a lens with diameter D = LL' (Fig. 2) and focal distance f. We know that the sun can be seen from the earth at an angle of 31'; hence the image of the sun ss' produced by the lens has a diameter of
$$d = 2f \operatorname{tg} \frac{31'}{2} = \frac{f}{110}.$$

The total luminous flux incident on the surface of the lens of an area of $\pi \frac{D^2}{4}$ is concentrated in the sun's image, the diameter of which is f/110 and the area $(\pi/4)(f/110)^2$.

The illumination E_1 of the sun's image, compared with E obtained directly is greater by the following factor:
$$\frac{\frac{\pi}{4} D^2}{\frac{\pi}{4} \left(\frac{f}{110}\right)^2} = \left(110 \frac{D}{f}\right)^2$$

Experience shows that a lens with a relative aperture D/f equal to 1/2 can set fire to bits of cork, dry wood shavings, etc. But this lens intensifies the sun's radiation by a factor of $(110 \cdot 1/2)^2 \approx 3000$. Thus radiation enabling us to rapidly burn up an object is 3000 times greater than the sun's radiation on the surface of the earth.

G_eneral case. So far we have only considered a case in which the sun is the source of light. Let us now study a more general case in which the source is an incandescent body accessible to us, for example, the crater of an electric arc
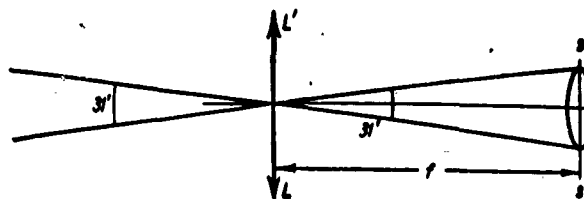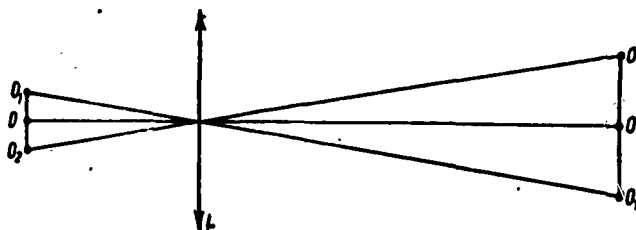
8

Fig. 2.   Image of the sun projected through a lens

Fig. 3.   Image of a light source at a finite distance.

Let O_1O_2 be the light source, L be the optical system represented by a lens for the sake of simplicity and $O'O'$ be the image of the source produced by the lens L. Let Φ be the power of the luminous flux incident on the lens. Our task is to determine this power, knowing the brightness of the source in different directions (a quantity which is directly measurable).

The most general case involves considerable difficulties in calculation and is of little practical interest. Let us therefore consider two particular cases which are very close to the light sources of greatest interest from the standpoint of application, namely:

1) when the light intensity is constant in all directions (for example in tungsten arc lights, known as point lights, excluding a very small area of shadow);

2) when the brightness of the source is constant over its surface and in all directions (the sun, the sun's image as produced by optical systems and arc craters belong to this category).

Let us further assume that the light source is located on the axis of the optical system, projecting its image onto the target, and that the system is symmetrical with respect to the axis (conditions always borne out in practice) (Fig. 4).
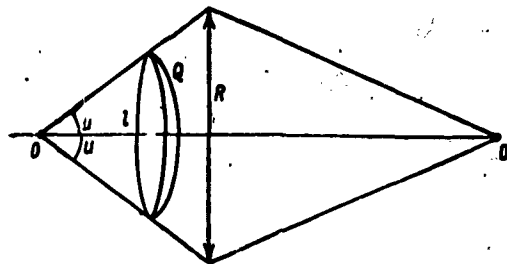
Fig. 4. Determining the luminous flux.

Let u be the angle formed by the axis of a ray passing through the edge of the lens. The solid angle at which the lens is visible from point O is determined by the area of the surface Q cut out by the cone of marginal rays on a sphere with a radius equal to unity, and by the center at the point O. This area is equal to $4\pi\sin^2\frac{u}{2}$.

In the first case the light intensity is constant in all directions and so the luminous flux emitted by the source and impinging upon the optical system is equal to the product of the light intensity I_0 and the solid angle, $i.e.,$

$$\Phi = 4\pi I_0 \sin^2 \frac{u}{2}. \tag{1}$$

In the second case the brightness B is constant; the intensity is not constant, and varies according to

$$I = Bs \cos u,$$

where s is the area of the source. Simple calculation (see any textbook on photometry) gives the following expression for the flux

$$\Phi = \pi Bs \sin^2 u. \tag{2}$$

It is interesting to compare Eqs. (1) and (2). If the source is fairly small, as u we can take the angle formed by the marginal ray passing through the center of the source towards the edge of the optical system. Furthermore, by considering the brightness to be constant over the entire source we get

$$Bs = I_0,$$

10

where I_0 is the intensity of the light in the direction of the axis of the system, and the second equation for $\tilde{\Phi}$ can be written in the form

$$\Phi = \pi I_0 \sin^2 u. \tag{2a}$$

We can see that at small angles u, when it can be assumed that $\sin u = u$ (the angle u is expressed in radians), both equations give the same result

$$\Phi = \pi I_0 u^2,$$

which can be further written in the form

$$\Phi = \pi I_0 \left(\frac{R}{l}\right)^2,$$

where R is the radius of the lens aperture in the optical system and l is the distance from the source. Or, still more simply.

$$\Phi = \frac{I_0 S}{l^2} = SE,$$

where S is the area of the lens and E is the illumination produced by the source on the lens surface. Thus, the flux is equal to the product of the lens area and its illumination, which is what should be expected since the flux is the product of the area and the illumination of it.

Nevertheless this case is of little interest to us, since the luminous flux cannot be very great when the lens is small in size.

Let us go on to consider the case of large angles u, which relates to searchlight mirrors (Fig. 5). Here Eqs. (1) and (2a) give different results; the equation for "point" sources, i.e., for sources with a constant light intensity, produces larger fluxes than the formula for sources conforming to Lambert's law (B is constant). For example, at u = 90°, we get for the former

$$\Phi = 2\pi I_0,$$

and for the latter

$$\Phi = \pi I_0,$$

i.e., half as much. In the second case a further increase in the angle u is of no advantage for practical purposes. By using complex optical systems, for example, searchlight mirrors, we can trap a considerable amount of the total power of the light source, as shown by Eqs. (1) and (2). It is indeed this _that_ THAT suggests the

possibility of concentrating a powerful flux on a distant object and setting it on fire.

But in concentrating the flux we encounter an insuperable difficulty - the dissipation of radiant energy.

Sec. 3. Dissipation of radiant energy

A powerful beam of light has *now* reached the optical system. All we need do is turn it into a parallel or slightly converging beam and the problem is solved. It might seem that nothing could be easier than obtain the desired result by means of a correctly calculated optical system. But at this point laws of a extremely universal nature come into play; these laws are corollaries of the law of *the* conservation of energy. They govern the distribution of radiant energy over large distances from the source.

The site of the greatest concentration of the rays passing through an optical system is the image of the light source*. Hence the area of the greatest concentration of rays emerging from the optical system is determined by the size of the image (Fig. 6).

The dimensions of the image can be calculated from geometrical optical equations. In the case in point we will apply the well-known Lagrange-Helmholtz, according to which

where n and n' are the refractive indices of the media in which the object and image are located, respectively; l and l' are the lengths of the object and image, u and u' are the angles at which the optical system is visible from the center of the object and the center of the image.

The Lagrange-Helmholtz law plays an important part in photometric calculations and is in *FACT* a modification of the principle of the conservation of energy. We should point out that the Lagrange-Helmholtz law holds for any optical system, no

---
*
Or else the optical system itself; but the latter is not of interest in the case in point.
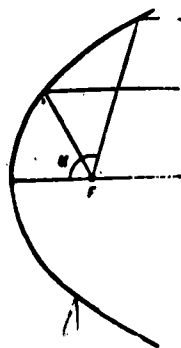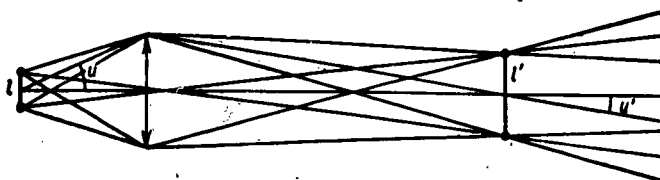
Fig. 5. Aperture angle of a parabolic mirror

Fig. 6. Structure of an image.

matter how many parts *IT CONTAINS* and no matter whether they are reflecting or refracting parts. There is no combination of optical systems which can invalidate the law, and this is why all hopes for an ideal concentration of radiant energy must collapse. Furthermore, the law even remains valid for generalized optical systems, for example, those comprising separate zones, such as lighthouse lenses (Fresnel lenses) or those employing refraction effects in air layers in which the refractive index varies, and also for "light lines" and beams of glass thread ("fibrous optics"), which have begun to be used in recent time. The proof of the Lagrange-Helmholtz law is given in the appendix.

Let us derive an expression for the illumination of an image produced by an optical system.

If we disregard energy losses in the system, the total power of the flux incident on the system reaches the plane of the image and is distributed over the latter. If we examine a fairly small element of the source, the brightness of the latter is constant over its area and the corresponding element of the image is therefore uniformly illuminated as well. Let us assume for the sake of simplicity that the element is circular in shape with radius r. Let r' be the radius of the *CIRCULAR* image . According to the Lagrange-Helmholtz law we get

$$nr \sin u = n'r' \sin u'.$$

Since in all practical cases n = n' = 1,

$$r \sin u = r' \sin u'.$$

Let Φ be the power of the incident flux. Since the outgoing flux Φ' has the same power, $\Phi' = \Phi$, and the illumination of the image element is then equal to

$$E' = \frac{\Phi'}{s'},$$

where s' is the area of the image, or

$$E' = \frac{\Phi}{\pi r'^2}.$$

Substituting for r' the value of the latter from the Lagrange-Helmholtz law, we get

$$E' = \frac{\Phi}{\pi r^2} \left(\frac{\sin u'}{\sin u} \right)^2. \tag{3}$$

It only remains to replace Φ by the formula for it in terms of the brightness of the source and the scope angle u.

Let us first consider the most important case in which the source conforms to Lambert's law. We then get from Eq. (2a)

$$\Phi = \pi I_0 \sin^2 u.$$

If we now express the light intensity I_0 in terms of B and area of the element πr^2, then

$$I_0 = B \pi r^2,$$

after which we find for Φ

$$\Phi = B \pi r^2 \cdot \pi \sin^2 u,$$

and for the illumination E'

$$E' = \pi B \sin^2 u'.$$

So far we have disregarded losses of light in the optical system. Let us use k (a number always less than unity) to designate the coefficient showing how much radiant energy is able to pass through the optical system. We then finally get for E'

$$E' = k \pi B \sin^2 u'.$$

$$\tag{4}$$

We should point out that Eq. (4) only holds in the case in which the entire cone of rays incident on the element of the image is filled with rays. We will see later why this point has to be stressed. Equation (4) can be written in a different way. On account of the great distance between the optical system and the target, the

angle u' is always small and we can therefore write

$$\sin u' = \frac{D}{2l},$$

where D is the diameter of the system*, and l is the distance between the system and the target. Equation (4) then takes the form

$$E' = \frac{kB}{l^2} \pi \frac{D^2}{4}. \tag{4a}$$

Let us use S to denote the area of the lens (in the general case this is the *exit pupil* of the lens), then

$$E' = \frac{kBS}{l^2}. \tag{5}$$

This equation is called the Mangen-Chikolev formula**.

In one sense Eq. (5) is more general than Eq. (4). S can be taken to mean the area of all parts of the optical system which when viewed from the illuminated target appear to shine. Thus, the Mangen-Chikolev equation enables us to consider the case in which several optical systems operate at the same time, although they may not constitute one unit, for example, a system consisting of separate mirrors (Fig. 7). But this system must satisfy the condition that all the mirrors should have the same purpose. For a mirror system of this kind Equation (4) is meaningless, since we do not know what is meant by the angle u'.

What are the implications of the Mangen-Chikolev equation? We can see that the illumination of any point on which rays coming from a light source with brightness

---

* In the general case it should be said that the diameter *is* the *exit pupil* of the optical system, i.e., the diameter of the real or imaginary diaphram which restricts the beam of rays emerging from the system.

** Mangen (1825 - 1885) was a French military engineer, known for his work on light-houses, searchlights and other long-range illumination devices.

Chikolev, V. N. (1845 - 1898) was a Russian military engineer who wrote a number of original works on illumination engineering. He was the first person to derive the equation given here, although this formula is known as the Mangen formula.
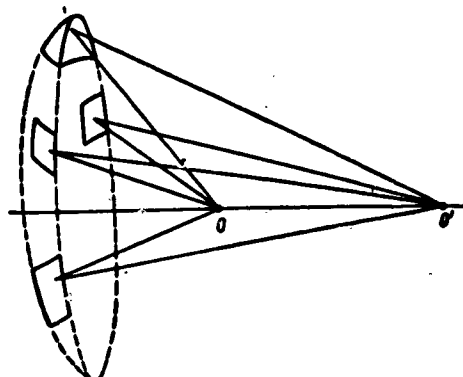
Fig. 7. Interpretation of the Mangen-Chikolev law for
a case in which the surface is not continuous.

B are focused, is determined in the following way. The optical system plays the part
of a new light source, observed directly from the point under consideration. The new
source has the same brightness as the original one (disregarding losses in the optical
system itself), but its area is equal to the area of the EXIT PUPIL of the optical
system.

Let us calculate the illumination produced at a distance of 1 km by a mirror
2 m in diameter, at the focus of which there is a high intensity arc with a brightness
of 100,000 stilbs. Disregarding losses, the illumination is (see Eq. (4a))

$$E' = \frac{100\,000\pi\,(100)^2}{(100\,000)^2} \cdot 10^4 \; \text{lux} = 3 \cdot 10^3 \; \text{lux}.$$

Despite the considerable power of both the source and the optical system, this
illumination is 30 times less than that produced by the sun on the surface of the
earth! Here we have not taken into account losses in the optical system or the
atmosphere.

Let us consider some more implications of the Mangen-Chikolev equation.
The illumination of a target does not depend on the size of the source.

An increase in the size of the source can only be brought about by enlarging the target, but not by increasing its illumination. The illumination of the target decreases in proportion to the square of the distance - which is quite natural.

Let us consider another case of a "point" source, i.e., a source with a constant light intensity in all directions. Let us return to Eq. (3)

$$E' = \frac{\Phi}{\pi r^2}\left(\frac{\sin u'}{\sin u}\right)^2 .$$

Instead of $\bar{\Phi}$ we will write down the expression for it IN Eq. (1)

$$E' = \frac{4\pi I_0 \sin^2 \frac{u}{2}}{\pi r^2} \cdot \frac{\sin^2 u'}{\sin^2 u} .$$

But since

$$\sin u = 2 \sin \frac{u}{2} \cos \frac{u}{2} ,$$

we get

$$E' = \pi \frac{i_0}{\pi r^2} \cdot \frac{\sin^2 u'}{\cos^2 \frac{u}{2}} .$$

And since $\frac{I_0}{\pi r^2}$ equals B,

$$E' = \frac{\pi B \sin^2 u'}{\cos^2 \frac{u}{2}} .$$

(6)

Equation (6) may seem to suggest the possibility of obtaining an infinite degree of illumination, by assuming u = 180°. But this is clearly contrary to common sense and IN FACT impossible. The expression for E' does not become infinite, but rather indeterminate, since both the numerator and denominator vanish. If we analyze this indeterminateness by means of the conventional methods of mathematical analysis, we easily see that

$$E' = 2\pi B \left(\frac{f}{l}\right)^2 ,$$

where f is the focal distance of the system, and l is the distance between it and the target. This value cannot be very great since the ratio f/l is always small. In practice there is no point in going further than 90 - 120° for the angle u, since the size of the mirror increases rapidly with this angle. At 135°, $\cos^2 \frac{u}{2} = 0.5$, and the illumination E' is double what it would be if the source radiated the same brightness, in conformity with Lambert's law.

Unfortunately, all point sources are much weaker than arcs (by a factor of 20 - 50 times), and are therefore unsuitable.

Sec. 4. General conclusions

The Mangen-Chikolev equation (5) shows that for a given light source and a given target distance, the <u>only</u> way of increasing the illumination of the target is to increase the transverse dimensions of the optical system. It is not difficult to calculate the size of the latter required to produce a fairly strong effect, let us say the rapid ignition of easily combustible objects such as paper, dry wood and so on.

In Section 1 we saw that a magnifying glass with a relative aperture of not less than 1:2 produces the desired effect, provided we use the sun as the source. Equation (4) shows that for the given source the angle u' determines the illumination of the target. Thus we come to the only possible conclusion - an optical system beaming the sun onto a target must be such that the diameter of the aperture is at least half the distance to the target; it is only in this case that u' has the same value as in the magnifying glass experiment.

The result of theoretical calculations is confirmed by experiment. In 1747 the French naturalist Buffon constructed a firing device consisting of 168 glass mirrors 6 inches long and 8 inches wide, arranged so that the gaps between them were not wider than 1 cm. The mirrors were attached to a common frame which turned in all directions.

In addition, each mirror had its own frame so that it could be rotated separately. Movement of the apparatus made it possible to concentrate all 168 images of the sun on one spot and set fire to objects at a considerable distance. It was possible to set fire to resinous pinewood planks at a distance of 158 feet. The focus, which was a circular image of the sun, was 16 inches in diameter; it took several minutes for the fire to get going.

It is easy to calculate that all Buffon's 168 mirrors were equal to one large mirror with an area of 5.9 m^2. It produced an illumination of 470 phots at a distance of 47 m, i.e., 36 times more than the illumination of the sun on the earth.

The flux attained 5 watts per cm^2 of surface and was able to raise the temperature

18

to 700°C. An equivalent mirror would be visible from the target at an angle u' = 3°.

For instantaneous ignition we require an illumination hundreds of times greater, hence u' must be 30° or more. If the range of the target is 1 km, for example, the diameter of the optical system has to be 500 m. There is no point in stressing how fantastically difficult a system of that kind would be; the possibility of constructing it and setting it in action in full view of the enemy is difficult to imagine.

But let us suppose that the enemy is not able to prevent the construction of a mirror 1 - 2 km in diameter, and that it is carried on in full view of THEM, let us say a few kilometes away. It would not be at all difficult to render the weapon harmless. All that need be done is to paint all easily combustible objects of value white or cover them with aluminum or some other foil AND any danger of a conflagration is prevented. White paints reflects 70 to 90% of the incident luminous energy, and the portion absorbed constitutes no danger.

As an illustration, we chose the sun as the light source; although this is possible in principle, it would require additional mobile systems comprising collosal mirrors in order to beam the light in the required direction.

All other convenient sources on the earth are less bright than the sun; to obtain any perceptible effect, the optical system would have to be even larger.

Sec. 5. Sources of heat energy

Leaving aside the sun, which is hardly suitable as a source in view of its movement through the sky, and other factors (for instance, the (possibility of at night or in cloudy weather), the best source of heat rays at the present time is a high intensity arc, the brightness of which, particularly in the red region of the spectrum (thermal region), is only slightly inferior to the sun.

Under laboratory conditions it is possible to produce brightness which is even greater, for example, by passing current through a thin metal wire, AND / MAKING the latter burn through with a flash, the brightness of which is many hundreds of times

19

greater than that of the sun. But these flashes only last for a few hundred thousandths of a second and the energy radiated by them is negligible.

In recent time great progress has been achieved in the development of pulse tubes, in which a flash is produced by instantaneous discharge of a battery of condensers (lasting several microseconds) through couples made of mercury or other elements. The brightness of the flashes is several hundred times greater than that of the sun. But since the duration of the flash is not more than a few microseconds and the effect is interrupted for comparatively long periods, the averaged brightness is low, and the sources cannot be considered important from the point of view of causing fires at a distance.

Last but not least, at the present time much brighter sources of light are known -/the atomic and hydrogen bomb. But the latter could hardly be used as light-sources for long-range combustion, and, furthermore, they perform the task without the need for optics, so we will not consider them here at all.

Naturally engineering is making big strides forward. Light sources now have such practical importance that a great deal of effort and money is being spent on improving them. Nevertheless, when we calculate the brightness required for the idea of long-range combustion to move beyond the realm of fantasy and become a practical possibility, we arrive at conclusions which inventors will find depressing. It can be assumed that an optical system of the searchlight type is the absolute limit from the point of view of size, since otherwise the camouflaging of a huge piece of apparatus becomes too complicated, especially when it is considered that the premises of the problem make it necessary for the apparatus to be visible on the enemy side (or else it will not work).

At this juncture we should turn out attention to a fact which might make it possible to increase the element of surprise, namely, the elimination of the DAZZLING effect of the apparatus during operation by placing in the path of the rays a filter which absorbs the visible rays, but lets through the most effective infra red portion of the luminous energy.

Assuming the diameter of the optical system to be 2 m and the range of the

target 1 km (provided the aforementioned illumination is produced on the target (p. 30)) we will have to intensify the brightness of the source by the square of the diameter ratio $(500/2)^2$, i.e., by a factor of $(250)^2$ or 60,000!

The writer would like to feel that he has been able to convince readers of the impossibility/of solving the problem of setting fire to distant objects by means *IN THE NEAR FUTURE* of optical systems . The methods of calculation set forth above are so general and so simple that to doubt them would be to doubt the fundamentals of nature.

Nevertheless, the more skeptical readers may still find several loopholes which leave hope that the general laws can be either got round or *THAT* considerable improvements can be brought about. The following section is intended for these readers.

Sec. 6. Limits of accuracy in calculating illumination

There is no doubt that the earlier derived equations are not exactly accurate. They almost always give an exaggerated illumination. For purposes of simplification we disregarded a number of factors causing energy losses. We made only very brief mention of losses occurring when rays pass through an optical system and through the atmosphere filling the space between the source and target. These losses may easily be as much as 40 - 60% or more of the total useful energy.

But let us look at other factors which in some way or other effect the validity of the target illumination calculation. We have only considered two types of sources - point sources and those conforming to the Lambert law. What about other sources? Is it possible to obtain more favorable results with particular emission laws? If we take B to be the maximum brightness of the source, then it is obvious that no distribution of source brightness can give higher values than those produced by point sources. Hence we should not strive for a particular distribution of brightness on all sides, but merely fairly bright sources.

The aberration sometimes gives rise to hope that the illumination produced by these systems can be increased. There is no point here in making a detailed analysis of aberration; basically speaking, it means that rays travelling from a point
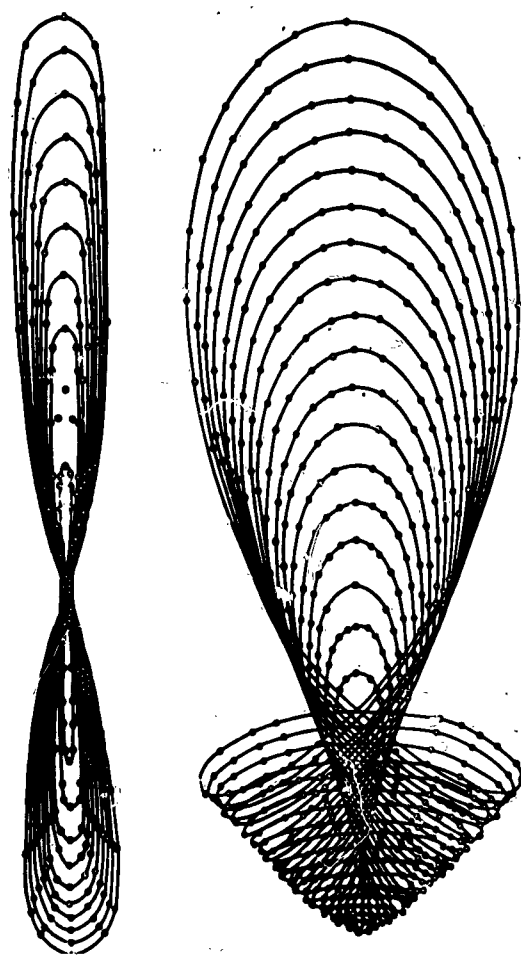
21

Fig. 8. Image of point when there are blur and astigmatism in different ratios.

on an object after emergence from an optical system do not gather strictly at one point, but rather in a slightly dispersed configuration. Examples of this are intersections by the focal plane, of light beams refracted by optical systems, which are emitted by points lying off the optical axes of the system when the latter is blurred or astigmatic: these are shown in Figure 8.

Aberrations are a vertible scourge for optical instruments. The need to counteract them had made the calculation of the instruments a complex, boring job and has given birth to a new, rather specialized branch of optics. But, as always, every harmful effect (even aberration) can under certain circumstances be turned into something useful.

Generally speaking, the presence of aberrations increases the dispersion of the rays conforming to the laws of geometrical optics and reduces the illumination as determined from the equations considered above. In certain cases the aberration caused redistribution of the energy in the plane of the image and may cause a certain concentration in the center through reduction at the edges.

We have an example of this compression in parabolic searchlights. Let AOB (Figure 9) be a section of the reflecting surface of a parabolic mirror, and let S be a point source of light in the focal plane at a distance 1 from the axis. Calculations show that the marginal rays SAA' and SBB' form a smaller angle with the axis after reflection from the mirror than does the central ray SOO', which is propagated in accordance with the laws of paraxial optics (the angle SOO_1 is equal to the angle FOO'). Despite this concentration of rays at the center of the image at infinity, however, the illumination in the center of the image, as shown by calculation, is not greater than it would have been in an ideal system.

There is one more phenomenon associated with the propagation of light which may raise the spirits of those seeking new ways of concentrating light rays. It is the defraction of rays by an optical system. We will come back to this phenomenon later; it is considered in detail in Chapter IV.

Suffice it to say at this point that the result of defraction is a certain

23
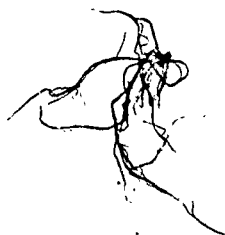
redistribution of luminous energy in the light beam, but of a different nature to the redistribution due to aberration. It may reduce the harmful effect of aberration, though it may not be in a position to improve the illumination derived by general equations. However, even this beneficial effect by defraction may only show up in the case of extremely small and therefore low-powered light source. In the existing sources, the dimensions of which are fairly large compared with the focal distance of the projecting system, the redistribution of the light is completely eliminated by the superimposition of images of individual points on the source.

A great number of physical phenomena, apart from those mentioned, may possibly be summoned to the investigator's assistance in order to solve the problem in question. But it must be taken into account that none of them can play anything but a subsidiary, side role; the most important phenomena for the given case are, of course, the propagation of luminous energy in accordance with the laws of optics and photometry. They, and they alone, can uniquely solve the problems facing us, and as we have seen, leave no hope of the possibility of using optical systems situated a long way from the target to bring about any great degree of heating in the latter.

It should be kept in mind that these conclusions follow directly from the general laws of physics, such as the principle of the conservation of energy; hence their reliability is not inferior to that of this indisputable principle. The time has not yet arrived for long-range combustion and we can only hope that it will never come since the need to cause conflagrations in peace time is unimaginable.
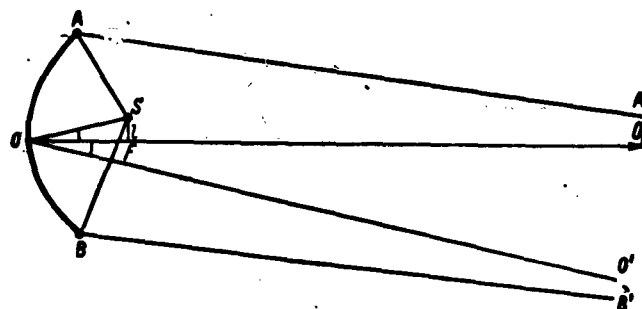

Fig. 9. Concentration of rays at center of parabolic mirror.

CHAPTER II

OPTICAL MYTHS

Section 1. Introduction

Specialists in optics often have occasion to deal with suggestions for the calculation or design of optical systems meeting certain requirements, the manufacture of which is beyond the possibilities of optics*. Hence among our scientific intelligentsia, even among recognized specialists, there is often a lack of clear understanding on what can be expected from optical systems. This lack of understanding is basically due to an erroneous impression, which can be formulated in the following way: optical systems enable a beam** of any structure to be converted into another beam of any, preordained structure.

It should be pointed out that the link between the given problem and this general assumption is so deeply concealed that both the poser of the problem and the specialist in optics called upon to solve it fail to observe it. The optical designer gropes for a solution, piles lens upon lens and prism upon prism, breaks up the beams and splits them asunder, wastes lots of time and finally comes to the conclusion that the problem is insoluble.

Let us consider in greater detail why the concept formulated above is wrong. Let us agree to take the term "beam" to mean the following (keeping in mind the fact we are dealing with applied optics and cases encountered in practice, we will keep to the narrower definition). A beam of rays is the sum total of the rays filling all cones, the apeces of which lie on the surface of the light source (or its image) and

---

* In this chapter we will not touch on a number of subtler problems involving the theory of aberration - a difficult branch of geometric optics. This will be dealt with in greater detail in Chapter IV, Section 9.

** It will be explained in more detail what the term "beam" means further on.

the common base of which is the ENTRANCE/of the optical system. This is different from the concept adopted in mathematics; in the given case the beam is an energy carrier and therefore possesses, as we have seen in Chapter I, a certain power (or flux), determined in the case of an elementary beam by the product

$$\Phi = Bs \cos \theta \cdot \omega. \tag{7}$$

Equation (7) will soon be needed. Furthermore, an important part is played by the Lagrange-Helmholtz law which we have already come across (page 24)

$$nl \sin u = n'l' \sin u',$$

and which in the cases of interest to us, where n = n', it takes the more simple form

$$l \sin u = l' \sin u'.$$

Let us transform this equation. Let us square both sides

$$l^2 \sin^2 u = l'^2 \sin^2 u',$$

where l^2 is the area of the square, the length of one side of which is 1, and l'^2 is the area of the square image of the former. But this equation can be generalised by writing it in the form

$$s \sin^2 u = s' \sin^2 u', \tag{8}$$

where s and s' are two adjoining areas. If the angles u and u' are small, $\sin^2 u$ and $\sin^2 u'$ can be replaced by the conjugate solid angles ω and ω', i.e., the solid angles bounding the conjugate beams. This formula can then be written in the simple form

$$s\omega = s'\omega'. \tag{9}$$

When we gave the proof of this law, it was assumed that l and l' related to the magnitude of the object and image. Furthermore, the proof might have given the impression that the Lagrange-Helmholtz law is only valid for conventional optical systems possessing an axis of symmetry and obeying the laws of paraxial optics; in other words, cone-shaped and similar-type surfaces are ruled out.

In actual fact, the Lagrange-Helmholtz law, inasmuch as it expresses the conservation of energy, is of more general importance and can be expressed as follows.

Let us imagine a beam of light delineates an area s on a plane perpendicular

to the beam axis, and that all the rays in the beam are bounded by the solid angle ω ; the products $n^{-2}s\omega$ is then constant <u>along the entire beam in any medium and in any system.</u> This principle, which can be considered as a generalisation of the Lagrange-Helmholz law, holds for small s and Φ . If the first and second media are identical, we can write down the equality

$$s\omega = s'\omega'. \tag{9a}$$

But as distinct from Eq. (9), s and s' in this case are not optically conjugate areas, and the optical system may be of any design.

Equations (7) and (9) prove the impossibility of those very alluring and, as many people think, easily attainable aims.

Sec. 2. "Parallel" Beams

What tremendous possibilities are promised by the very existence of parallel beams even though they may carry a very small amount of energy! By means of telescopic systems (for instance, astronomical telescopes) it should be possible to concentrate them into needle-sharp beams and thereby solve the problem of combustion from a distance. A simplified optical system of this kind is shown in Fig. 10, in which, by means of three telescopic systems a parallel beam can be transformed into a "needle". This is how many inventors solve the problem (in principle), among others the hero in A. Tolstoy's well-known adventure story "The Hyperboloid of Engineer Garin". A great deal is said about parallel rays in books on optics; the rays play a prominent part in explanations of the properties of optical systems to such a degree that the existence of such beams under actual conditions may not be questioned at first sight.

Where, then, is the error?

There are two errors, and they are fundamental ones: first, parallel rays do not exist, second, even if they did, they could not carry any energy.
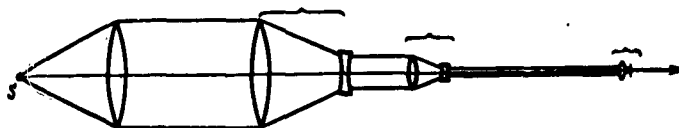
SUBMITTED

Fig. 10. One of the DESIGNS/ for "long-range combustion machines".

We all know that to obtain a parallel beam, we have to place THE luminous point in the focus of an optical system. But to do this, we have to observe three conditions, which are not REALIZABLE to an equal extent.

1. The use of a luminous point without dimensions as the source. Such sources only exist in textbooks, though, admittedly, to a prolific extent; they are required for simplification of calculations. Under actual conditions there can be no energy radiator which does not possess definite dimensions, although they may be very small.

2. The existence of ideal optical systems, i.e., those which do not exhibit aberration. There are no such systems, nor can they ever be manufactured.

3. The absence of diffraction. Diffraction cannot be completely eliminated under any circumstances.

Thus, the very existence of parallel beams is totally impossible. To DISPEL in readers of all final doubt, let us ask one further question: is it possible to compensate for the finite dimensions of the energy source by means of the abberations, or by utilizing diffraction, in order to restore the parallel beam? It is easy to show that this compensation is not possible. If the point source diffuses the light to a slight extent on account of aberration and diffraction, finite dimensions of the source can only increase it, but cannot in any way reduce it.

Let us now prove that had it even been possible to produce a parallel beam of rays, the beam could not have carried any energy.

At first sight this may seem odd since the analogy of a stream of liquid flowing through a cylindrical tube and possessing energy inevitably comes to mind. But this is merely a play on words, and not an analogy. A stream of light is determined by $^Eq.$ (7) or modifications of it. If the rays in the beam are parallel, then 1) $s = 0$ and 2) $\omega = 0$. Consequently, $\Phi = 0$ and the power of the beam is equal to zero. Thus, parallel beams are pure fiction.

Sec. 3. "Concentrating Cone"

When designing different instruments, in particular, recording devices, designers have to solve the problem of the maximum possible concentration of luminous

or heat energy. After numerous attempts it is found that no combination of the conventional optical systems produces the required effect. They then resort to optical systems of a special kind which do not produce images. The distinguishing feature of the systems is the fact that there are either no planes at all tangential to the surface of rotation on the axis of the system, or there is an infinite number of them; an example of this system is the reflecting cone (Fig. 11), whose axis of symmetry 00' is the axis of the system as well. A beam of rays impinging upon the base of the cone 0 undergoes multiple reflection and eventually emerges through the opening 0', which may be as small as required; this system indeed solves the problem, or at least it does at first sight.

We need only trace the course of the ray, however, to see (Fig. 12) that with every reflection the ray slows down, after which it begins to turn round and finally emerges through the base 0. A very small number of rays reach 0' and emerge from it; here, the smaller the opening, the less the power of the emerging ray. The equation determining the power remains valid, of course, as can be shown in greater detail, though the operation is rather laborious.

It should be pointed out that the use of "concentrating cones" is advisable in certain cases, for example, whenever THEIR use does not conflict with the generalized Lagrange-Helmholz law. Let us assume that the area of the opening 0 is equal to s and that the area of the opening 0' is equal to s'. Let θ be the plane angle of the solid angle ω, and let θ' be the plane angle of the solid angle ω' of the beam emerging from the cone and reaching the area s'. We get

$$s\sin^2\theta = s'\sin^2\theta'.$$

Let the plane angle at the apex of the cone be equal to β. Unless θ' is larger than β, the beam reaches the area s'. Hence the condition ensuring, when fulfilled, that the TOTAL FLUX which enters the cone also emerges from it, can be written as follows

$$\sin^2\theta < \frac{s'\sin^2\beta}{s}, \quad \text{или} \quad \sin\theta < \sqrt{\frac{s'}{s}}\sin\beta.$$

It should be pointed out that the problem considered here is in fact a partial case of a more general problem, namely: the passage of a certain amount of

luminous energy FROM a given light source through an opening of given size when there is an arbitrary optical system in the path of the beam. This problem will be considered in greater detail later on (Sec. 7 in this chapter); here we will merely say that the solution of all problems of this type is based on Eq. (9):

$$s\omega = s'\omega',$$

i.e., the product of the source area (or its image) and the solid angle (divergence of beam radiated by this area or forming the image) is a constant value. The smaller the image, the greater the solid angle.
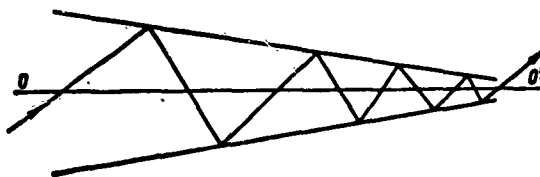
Fig. 11. "Concentrating cone".

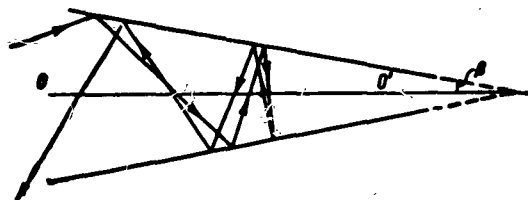Fig. 12. Path of the ray inside a meridional cross-section of the "Concentrating cone".

Sec. 4. Problem of optical intensification in all directions

Let us assume that a body (a ship or an airplane) is sending out visible or invisible (infra-red) rays. It would be of great interest to be able to detect these rays, no matter in which direction the emitter is radiating them with respect to the observer.

Since the emitter is a long distance away, the signals from it are weak and have to be amplified. This problem is neatly solved by an optical system of the simple spyglass type (Fig. 13). Let O_1 be the objective, O_2 be the eyepiece and P be the EXIT PUPIL of the system, i.e., the image of the objective produced by the eyepiece. The receiving device (for example, a photo cell or a thermo-element) is positioned so that the light-sensitive (heat-sensitive) layer coincides with the PUPIL . Let us assume that the telescopic system $O_1 O_2$ produces an angular enlargement $\dfrac{u'}{u} = \gamma$. As is well known, in telescopic systems, the linear enlargement is constant and is the reciprocal of the angular enlargement. Hence the diameter of the EXIT PUPIL is γ smaller than that of the objective. All the energy striking the objective O_1, next strikes the PUPIL P, but is concentrated over an area

γ^2 times smaller than the area of the objective. If we disregard losses inside the optical system, it can be said that the receiving element gets γ^2 times more energy that it would do, had there been no optical system. It is conventional to say that the latter amplifie $Y = \gamma^2$ times. This is a rare case in which an optical system proves useful when not being used for its immediate purpose. However, having been put on our guard by previous experience, we feel that we will certainly have to pay in some way or other for the benefit derived. The solution to this problem must obviously be sought in the properties of telescopic systems, associated with enlargement.

In our case everything that we have gained in amplification is due to another, highly important property of the system, namely, the angle of the field of vision. This is a corollary of the Lagrange-Helmholz law, the law restricting the possibilities of optical systems in relation to the freedom of conversion of light beams. Without amplification the acceptance angle (the angle of field of vision) of the receiving element, i.e., the angle at the axis, inside which the signals act on the receiving element, is very large and theoretically close to 180°, although in practice it may be much less.

As soon as we put an optical amplifier in front of the receiving element,

the angle is greatly reduced. The angle formed by the axis of rays emerging from the eyepiece can not be greater than 90°, even in theory; in practice there have not been so far any eyepieces in which the angle was greater than 45°, and highly complex eyepieces greatly absorb light, to boot. Using simpler means, we can attain angles of 30 or 35°. Let us use 2u to designate the angle of field of vision of the optical system, i.e., the angle formed by the axis of rays passing through the optical system. It is determined by the Lagrange-Helmholz law

$$l' \sin u' = l \sin u$$

or

$$\frac{\sin u}{\sin u'} = \frac{l'}{l} = \frac{1}{\gamma} = \frac{1}{\sqrt{\gamma}} ;$$

from which

$$\sin u = \frac{\sin u'}{\sqrt{\gamma}} .$$

The greatest possible angle u_{max} is obtained whenever u' = 90° and sin u'=1. Hence we get

$$\sin u_{max} \leqslant \frac{1}{\sqrt{\gamma}} .$$

But in practice we should assume u' = 35°, and then

$$\sin u_{max} = \frac{0,7}{\sqrt{\gamma}} .$$

For example, when amplifying γ = 100, we get sin u_{max} = 0.07 and $2u_{max}$ = 8°.

The requirement that the optical system should exhibit both large amplification and a large field of vision at the same time (in one receiving device) is absolutely out of the question. No matter how cleverly we JUGGLE WITH (the different optical parts, it is no use. Even attempts to make the optical system work "in parallel", i.e., independently, but in such a way that the radiant energy is directed onto the same receiver, are useless. It is easy to prove that the optical systems here darken each other. The task is beyond the power of optical systems.
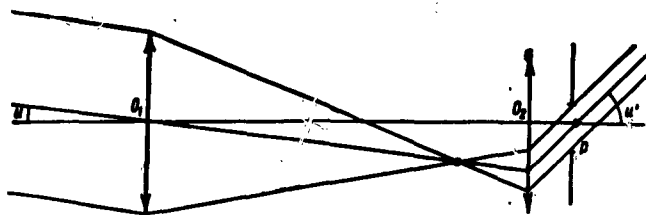
Fig. 13. Path of rays in a telescopic system.

Sec. 5. Photographic objective with relative aperture
greater than 1:0.5

Despite the fact that we are not dealing in this chapter with the restrictions imposed on optical systems by aberrations, which would distract us from the general principles of optics and leave us to flounder in the theory of calculations, we can not pass over a restriction imposed upon the speed of photographic and projection lenses by nature itself. It is associated with the theory of aberration by means of a simple, very general relationship producing an extremely important and evidently little-known result: the relative aperture of photographic and projector lenses cannot be greater than 1:0.5.

This follows from the so-called sine condition, which must be satisfied if we are to obtain a good image in the middle of the photographs, or screen onto which the picture is projected. The condition is written as follows:

$$\sin u = \frac{h}{f},$$

where h is the height at which the ray intersects with the ENTRANCE PUPIL of the objective, u is the angle between the ray and the axis after refraction (tan u $\backsim$ sin u on account of the smallness of the angle u), and f is the focal distance of the objective (Fig. 14).

Since sin u cannot be greater than unity, h $\geq$ f, and since the relative aperture is the ratio 2h/f, the sine condition implies the above-formulated premise. In certain instances this limit is practically attained.

There is, however, a method of going even further; by using SPECIAL fluids, into which the light-sensitive material is immersed. Here the relative aperture is increased n times, where n is the refractive index of the immersion fluid. It should

be pointed out that the objective should be calculated with specific consideration for the effect of immersion. Immersion objectives are now used in spectroscopy.

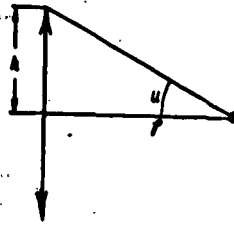To avoid confusion, we should explain the situation further.

Fig. 14. Sine condition

The impossibility of making objectives with a relative aperture greater than 1:0.5, (when the light-sensitive layer is in the air), only applies to objectives which have been corrected for aberration at non-axial points - such as all photographic lenses. But there is a whole crowd of high-grade optical systems with a much greater relative aperature, for example 1:1/3 or 3:1; 5:1 or even larger. Examples of these are parabolic mirrors USED IN searchlights, Fresnel lenses for lighthouses and certain condenser systems. These systems only require that the parallel beams are gathered strictly at one point, or, conversely, that the rays proceeding from a point source should emerge as a parallel beam. What happens to neighboring points is of no practical interest. Indeed, it can be proved that all the systems enumerated above exhibit tremendous coma and/ARE absolutely useless for photographing objects when the angular dimensions are to any extent different from zero.

Thus, the restriction which we encounter here is of a special nature. It is due to the need to represent correctly a small area, and not a point. This condition in fact determines the upper limit of the relative aperature.

Sec. 6. Conversion of diffused light into beamed light

In most OF THE obvious spheres of application of optics we deal with light sources emitting diffused light. If these sources are weak, we naturally try to alter the radiation in such a way as to increase it in one direction and reduce it in another. Such sources are the fluorescent screens used in television, fluoroscopy and radiography. It is just as tempting to increase the brightness of objects projected

onto a screen by episcopic devices whih the objects are illuminated by outside
sources and THE/ REFLECTION IS to some extent diffuse. , and therefore low brightness.

If it were possible to gather diffused light with an optical system and beam it at a certain solid angle, the problem would be solved. And it has indeed been suggested on many occasions that the diffusely-radiating surface should be covered with a network of microscopic lenses in the hope that the latter would gather the dispersed rays into a narrow bright beam.

Proposals of this kind are natural and understandable. On the one hand, they are due to analogy with certain screens which do/exhibit directed brightness, i.e., which reflect MOST OF THE incident light energy within a comparatively narrow solid angle, and, on the other hand, as in other inventions in the field of optics, a certain part has been played by the adjective "converging" mutually associated with the word "lens" and giving a completely wrong idea of its effect. The action of a lens of this kind is usually represented as shown in Fig. 15a and gives the impression that after refraction the rays emerge in narrow "directed" beams.

Let us define the words "ideally diffusing" (scattering) and "directed" more accurately, since there is certain confusion on this score.

An ideally diffusing source is one in which each point obeys Lambert's law, i.e., it radiates with equal brightness in all directions; furthermore, the brightness on all points on the surface is the same. If the surface of the source does not obey this law, the beam is directed. The more rapid the fall in brightness as the angle between the direction of the ray and the normal to the surface at the given point increases, so the greater the directivity.

In the problem of directivity of beams, a highly important, if not the/principal feature is the size of the source and its distance from the point S with respect to which the directivity is DETERMINED

If the source is small and the point S is far away from it, the source is a point source of light with respect to the point, and the beam is directed to a maximum extent since an extremely narrow beam emerges from the point S.
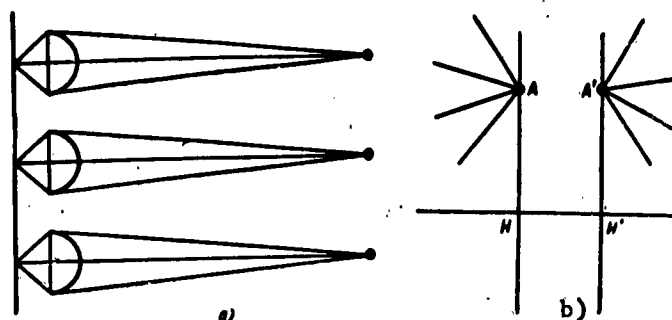
Fig. 15. Conversion of diffused into directed light.

If the area of the source is infinitely large - let us assume for the sake of the example, that the source is plane - the entire space on one side of it is ideally diffusing, since every point in this space is impinged upon by rays from all points ^{at} the source with equal brightness, and the area becomes an ideally diffusing source. All points in the space are situated on a plane parallel to the source and are subjected to the same conditions; hence any plane or any surface are ideally diffusing surfaces.

Let us now consider the effect of an optical system on the distribution of brightness in rays and on the directivity of the rays.

Let H and H' be the principal planes of an optical system (Fig. 15b). The arbitrary point A on the principal plane H is impinged upon by rays from all source points, hence from all directions. For every point A there is a point A' which by mapping point A with linear and angular enlargements equal to unity, radiates in exactly the same way as point A; the same thing happens to all the points on plane H. The plane H' constitutes the same ideally diffusing light source as the plane H. Thus, the presence of the optical system has in no way changed the structure of the radiation from the light source; it merely weakens the brightness of the beams, since some of the luminous energy is reflected and some absorbed.

This argument is valid, no matter how complex the optical system, since it is based on properties in common to all optical systems. This is the case with ideally diffused light when the source is of infinitely large dimensions. But it is of very little interest; much more important - since it crops up so frequently ● is the case of finite sources. But it can easily be shown that the latter case reduces to the former case in practice.

If the source is small in size, and the optical system is a long way away from it, only a small amount of energy radiated by the source reaches the optical system, while most of it is scattered in all directions. Hence we are only interested in the case in which the system is in direct proximity to the source.

If the ENTRANCE PUPIL of the optical system is much larger than the source, so that most of the energy radiated by the source impinges upon it, it is possible to obtain a direct beam. Here the optical system acts as a microscope objective with a large aperture, and at a large distance from the source we can obtain a considerably enlarged image of it, radiating over a comparatively small solid angle. But the solution of the problem is hardly of practical interest. We are thwarted here by the tremendous dimensions required for the optical system.

If the system is small in size, however, and is very close to the source, there is a repitition, with slight variation, of what occurred in the case of the infinitely large source.

Hence, under no circumstances can diffused light be turned into directed light. Any analogy with special movie screens reflecting a narrow beam of light is compltely superficial, since it is very narrow beams from the objective of the projecting equipment at a long distance from the screen which strikes the latter. These beams are of an acutely directed nature and we need To DO IS TO KEEP THEM THAT WAY ; this can be done easily in many different ways.

We will come back to this problem in Chapter III.

Sec. 7. The passage of a luminous flux through a narrow opening

Just as the Clausius law, the Lagrange-Helmholz law can also be called a constant flux law, and in that form it is in fact the law of the conservation of energy expressed in terms of optical characteristics.

Despite its simplicity and obviousness, it is sometimes disguised to such an extent that the greatest experts cease to recognize it. An example of this is the attempt to solve the OPTICAL/problem of the passage of as much luminous energy as possible through a given opening.

For example, let ABC (Fig. 16) be the section of a solid of revolution with

axis of symmetry OO'. A luminous flux is able to pass through the orifice with diameter CC'.

What is the maximum flux which can pass through the given solid?

It is determined by the formula

$$\Phi = \pi B \frac{\overline{CC'}^2}{4} u^2, \tag{10}$$

i.e., it is equal to the product of the area of the diameter CC' and the source brightness and the square of the aperture angle u. On account of the smallness of the opening CC', the aperture angle may be replaced by half the linear angle between AC and A'C'. It is easy to prove that no matter where the image of the light source *may lie*, a greater flux than the one given by Eq. (10) cannot be obtained. Thus, the maximum flux is totally restricted by the bounds of the beam and the brightness of the source, and given the dimensions of the free space and brightness of the source, it is absolutely impossible to go beyond the LIMITS SET by the equation. In order to obtain this maximum flux, we have to ensure, by means of a condenser, that the neck CC' is filled with the image of the light source, and we also have to ensure THAT THERE IS AN aperture angle u.

Another problem of illumination, the solution of which is of great practical importance, can be formulated in the following way. Let a source of light S_1 (Fig. 17), which for the sake of simplicity we will take to be circular, radiate a flux at a solid angle with aperture u_1. We are required, using an optical system, to pass the flux in toto through the opening S_2 (which can, among other things, be regarded as the working area of the receiver).
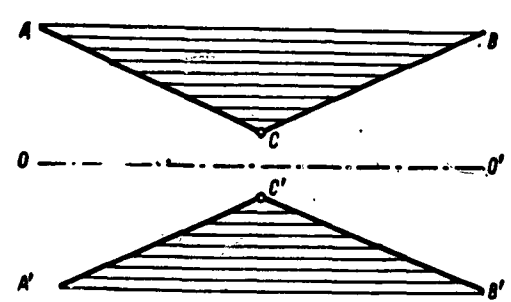
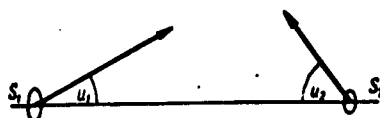Fig. 16. Passage of luminous flux through a solid of revolution

Fig. 17. Passage of a luminous flux through a narrow opening.

Most specialists encountering this problem for the first time gain the impression that a fairly complex system of mirrors and lenses will always enable them to "pass" any flux through S_2, and that the only difficulty is to construct an optical system which traps all the rays coming from S_1 and beams them through S_2. At first sight it may seem that there can certainly be such a system, although it may be a rather complicated one.

Everyone is fully acquainted with the Lagrange-Helmholz law, but many people when remembering the elementary proof of it given in most textbooks on geometric optics, assume that the law holds for centered, regular optical systems, and for conjugate planes as well. Hence there is some hope left that the law can be got round. This error would probably not occur, if it were stressed in the textbooks that the Lagrange-Helmholz law, as pointed out, is merely the principle of a constant flux, the universality and INVIOLABILITY of which is obvious to everyone.

It is easy to prove that the only luminous flux which an be passed through a preset opening is one which does not exceed a certain size.

For the sake of simplicity we will assume that the source obeys Lambert's law, i.e., we assume that the brightness is constant in all directions. The flux radiated by the source S_1 and the area s_1 is determined by the equation

$$\Phi = Bs_1 \sin^2 u_1.$$

Assuming that the optical system projecting the source S_1 onto the opening S_2, whose area is equal to s_2, does not causes losses, we find the following for the flux passing through S_2

$$\Phi = \pi Bs_2 \sin^2 u_2.$$

These equations give us

$$\sin u_2 = \sqrt{\frac{s_1}{s_2}} \sin u_1$$

If s_2 is such that

$$\sqrt{\frac{s_1}{s_2}}\sin u_1 > 1,$$

then the angle u_2 does not exist, and this means that it is not possible to pass the entire flux radiated by the source S_1 through the opening S_2. We can only pass some of the flux through; and the amount of it is determined by the condition

$$\sqrt{\frac{s_1}{s_2}}\sin u_1 < 1.$$

Sec. 8. The "fire risk" optical and glass parts

As we have often observed, people attribute amazing properties to optical systems, which the latter do not in fact possess. The potentialities of optical systems in this respect are often greatly over-estimated.

It is a fairly widely held opinion that optical systems, and further, a number of every-day household objects such as bottles and decanters, which only resemble optical systems to a very slight extent, may cause fires and explosions. This prejudice against optical instruments andother glass items is particularly widespread among those who engage professionally in fire fighting and keeping an eye on the observance of fire precautions. The tendency to exaggerate the "fire hazard" of optical instruments sometimes results in excessive precautions involving pointless expenditure and needless measures cramping freedom of action.*)

---

*) Illustration of this is the following extract from a widely used textbook on fire fighting (a translation from the German original by Dr. Schwartz "Conflagrations and Explosions").

"Any rays concentrated by a lens at one point may cause a fire within a short time (several minutes). A concentration of rays of the sun may be produced by any glass objects acting as lenses, for example, bottles, globes, empty and full vessels, measuring gauges, spectacle lenses, optical and camera lenses, glass tiles, or any glass objects containing air bubbles.

Glass tiles containing air bubbles are particularly dangerous. If used for the roof of barns or storehouses they may cause hay, straw or wood to burst into flame".

To gain a correct idea of the "fire risk" of optical instruments, we need only remember some of the information contained in the previous chapter (p. 16). Ignition of the most inflammable substances requires a concentration of sunlight by a factor of 50 to 100 or more. This is a large concentration and cannot be produced "by itself". It cannot be produced by just any lens. A relative aperture of the order of 1/10 or more is required. At the same time the diameter of the lens (or its focal distance) has to be such that the sun's image is fairly large - several millimeters in diameter, i.e., the focal distance has to be on the order of 20 cm or more. However, it should be pointed out that even at such high concentrations of solar energy (of the order of thousands or tens of thousands or more), when the diameter of the image is sufficient to ignite the object, it is much more often smoldering RATHER THAN COMBUSTION WHICH/ occurs, since the neighboring areas of the heated part cool it and prevent it bursting into flame.

But in order for the concentration of luminous energy needed for combustion to occur, the distance between the lens and the combustible object should be equal to the focal distance of the lens.

Thus three conditions have to be satisfied all at the same time: the presence of a positive lens with a relative aperture exceeding a certain lower limit; a set distance between the lens and the object and a fairly high degree of inflammability in the latter. The probability of all three conditions BEING satisfied at the same time is so small that it is virtually nil (although nott exactly nil).

As regards the "fire-raising" properties of bottles, decanters, etc., the "concentrating" power in them is negligibly small.

Bottles filled with liquid are similar in action to cylindrical lenses condensing the solar energy in one direction (in a band), rather than a point or, more exact, a circle, as is normal in lenses. Hence the illumination caused by large bottles (which must be full since empty ones do not produce any concentration for practical purposes) is considerably less than the illumination produced by the effect of a solid glass globe with a diameter equal to that of the bottle.

Fig. 18 shows the path of rays through a glass cylinder 200 mm in diameter with a refractive index of 1.5 (it should be noted that most liquids PRODUCE even less illumination on account of a lower refractive index). The width of the light band at the narrowest point AB is about 20 mm, and since the diameter of the cylinder is 200 mm, it can be considered that the cylindrical lens amplifies the illumination approximately 10 times, which is a very weak and quite harmless "concentration". Admittedly, the light is distributed unevenly over the bAND and if we take the angular diameter of the sun into account, it can be calculated that at certain points the concentration is somewhat larger, though it still does not exceed 15 - 20, AND THIS is a theoretical figure anyway.

In practice, because of absorption of the medium, scatter and other factors, we can consider that the degree of concentration never exceeds 8 - 12, which is absolutely harmless. A vessel shaped like a globe and filled with water may produce a higher concentration - an average of 50 - and at certain points as much as 100 times; this concentration is much more dangerous, but the great losses due to absorption by the glass greatly reduce it.

Empty bottles and spherical decanters do not concentrate light to this extent; the flashes occurring when sunlight impinges on these objects are very weak and do not present any danger. Small bubbles in the glass can only cause diffusion, and diffused light has still less of a heating effect than the sunlight itself. If there are large bubbles   , they act as negative lenses. This is easy to understand if we follow the course of rays in a lens of this kind (Fig. 19). (It must be taken into account that the refractive index of the glass is greater than THAT OF the air). Instead of (A) concentration of solar energy, WE FIND THAT (the latter is diffused.

Nevertheless, on the basis of this we should not jump to the conclusion that optical systems are completely safe. On the contrary, if the occasion arises, we can use them to produce exceptional results, unobtainable in other ways in the high temperature range.
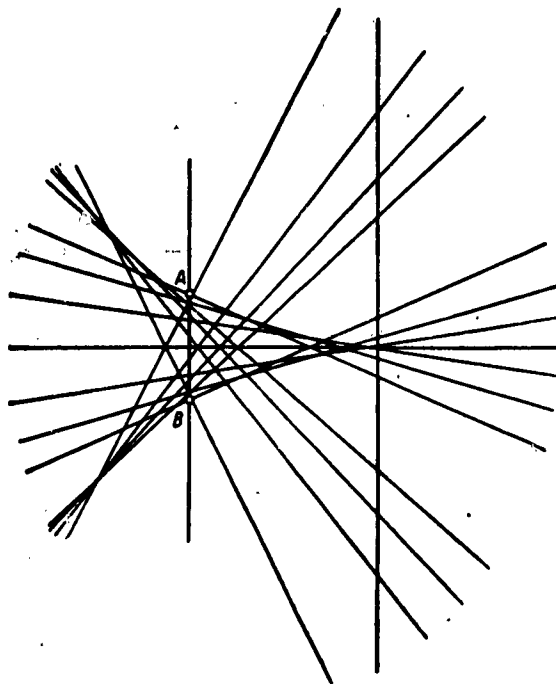
Fig. 18. Path of rays through a section of a glass cylinder.

Using   large paraboloid mirrors 2 m in diameter with a focal distance of 0.5 m, (the kind employed in set lighting) we can concentrate solar energy by a factor of 160,000, which is equivalent to an energy density of 3000 cal/cm^2. If a small stove, the volume of which is not more than several cubic centimeters, is placed at the focus of this mirror, and maximum use is made of the energy obtained by special devices, we can raise the temperature of this tiny stove to 3000°C. At this temperature there are a number of chemical reactions which cannot be produced in any other way; the highest melting metals can be brought to melting point this way.

It should be pointed out that a great part has been played in development of such solar furnaces by M. V. Lomonosov. It was he who thought out a clever and original design for a furnace in which the sun's rays heat it up on all sides by a special arrangement of plane mirrors and lenses.

At the present time the science of helio-installations is being rapidly developed; these are installations which utilize solar energy for a variety of purposes. Among them are ones in which sunlight is used in a strongly concentrated form, for example solar steam boilers. Here the rays are concentrated through

reflection of sunlight from large mirrors, the surface of which is calculated in such a way as to utilise the solar energy to a maximum extent.
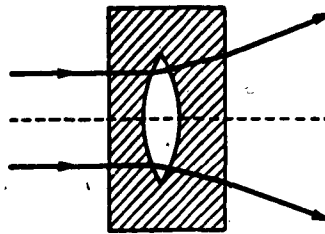

Fig. 19. Path of rays through an air-bubble in glass

Sec. 9. "Amplifying" the illumination

A number of inventors have at different times suggested a method for amplifying the illumination of an area S by means of the device shown in Fig. 20. Two conical surfaces, the axis AB of which coincides with the axis of symmetry of the objective OO', beam the parallel rays of light P and P_1 ONTO THE *OBJECTIVE* through the ring LL_1MM_1. and the rays are condensed in the focal plane FF_1 of the objective.

The area of the ring through which the rays pass is $\pi(r_1^2 - r_2^2)$ where where r_1 and r_2 are the outside and inside radii of the ring. But the area of the ENTRANCE PUPIL of the objective is πr^2, where r is half the diameter OO'. But $r_1 - r_2 = r$, hence,

$$\pi(r_1^2 - r_2^2) = \pi(r_1 - r_2)(r_1 + r_2) = \pi r(r_1 + r_2);$$

in this way the area of the ring is greater than the area of the ENTRANCE PUPIL by a factor of $(r_1 + r_2)/r$. This ratio can be as large as required. On the other hand, all rays incident on the ring are concentrated on the objective. On first sight it may appear that the illumination of the focal plane FF_1 becomes $(r_1 + r_2)/r$ times larger than the plane produced by the objective itself.

It is easy to check whether the rays lying in the plane of the drawing and forming the angle $\mathcal{N}$ with the axis AB gather at the same spot after passing through the ring as the rays directly striking the objective at the same angle to the axis. In other words, for plane beams of rays lying in the plane of the diagram, there is

concentration, provided the concept can be applied to a linear element (and not to an element of area). This property of the systems described has led many an inventor to the conclusion that by means of the optical device described it would be possible to intensify the illumination at the focal point any number of times. According to the illumination equation given on p. 26, however,

$$E = \pi B \sin^2 u,$$

where u is the aperture angle of the objective OO'. Consequently, the illumination cannot be increased by the above-described device. How can we explain the apparent contradiction?
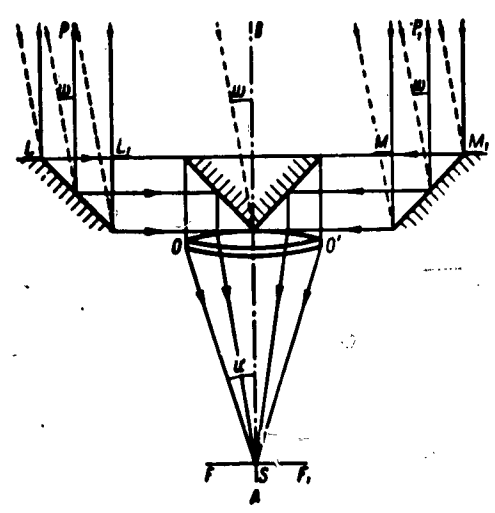
Fig. 20. Optical device for "amplifying" the illumination.

The answer is the behavior of the rays which do not lie in the plane of the diagram.

Calculation of the path of these rays shows that they do not gather at one point, like the rays in the plane of the drawing, but are scattered. The scattering increases with the distance between the conical surfaces. Hence there cannot be any concentration of the rays and the general equation for E given above still holds.

It can be further pointed out that the system in question does not produce any image, or at least as far as the representation of point by point is concerned.

As can be seen from the foregoing, inventors of the "illumination amplifier" make a mistake in that they only consider rays whose behavior is easiest to determine, and without foundation draw general conclusions from the results of a

particular case.

The error made by inventors who put forward the "amplifying" unit described is very like the one which will be considered further on (Ch. III, Sec. 3), since in general cases the error is due to insufficient consideration of the sagittal rays.

Sec. 10. Restoring sight in cases of retinal injury

Among the eye troubles suffered by elderly people, injury to the retina is very common; as a result of hemorrhage or other factors, part of the retina ceases to react to stimulation by light. Thus the working area of the retina is decreased compared with the normal. The DECREASE causes a reduction in the field of vision of the patient's ocular system and obviously causes a great deal of inconvenience - difficulty in getting about, working, and so on. The inconvenience is further aggravated by the fact that in many cases there is damage to the retina of both eyes at the same time. Is it possible to restore the partially destroyed vision by means of an optical system?

Let us recall the structure of the retina. The outside layer, which receives luminous energy, has a discrete cellular structure. Slightly simplifying the situation, we can say that the layer consists of a considerable number of separate cell-receivers (rods and cones), the majority of which are SUPPLIED with nerve endings. In the middle of the retina every light-sensitive cell has a nerve ending, but in the peripheral areas only groups of cells have them.

Thus, even if we assume that every cell could send a separate signal to the brain, the number of such signals could not be greater than the number of light-sensitive cells. Consequently, if part of the retina is destroyed, the number of working cells is reduced and there is bound to be some loss of information (see Ch. V, Sec. 8).

Let us assume for the sake of simplicity that the retina is shaped like a circle with diameter a. Let us assume that as a result of illness the diameter of the working area of the retina is reduced to a_1. By means of an optical system we can project the image of the same field of vision that existed before onto a new area of the retina. In order to do this all we need is to place special, so-called

telescopic glasses in front of the eyes and these will produce an image of distant objects on the retina with slight magnification (greater or less than unity). Telescopic spectacles are usually used to magnify AN image, and this to some extent offsets the lack of acuity characteristic of persons suffering from damaged retinas. Here, however, the field of vision is reduced, as is implied by the Lagrange-Helmholtz law.

In the case in point, we would have to wear telescopic spectacles with a magnification $\dfrac{a_1}{a_2}$ 1. Indeed, if we used f to designate the leading focal distance of the optical system "telescopic spectacles - eyes" and 2w to be the angle of field of vision of a normal eye, then the diameter of the working area of the retina is determined by the equation

$$a = 2f \operatorname{tg} w.$$

For a damaged retina we also get

$$a_1 = 2f_1 \operatorname{tg} w,$$

from which it follows that

$$\frac{a}{a_1} = \frac{f}{f_1}.$$

But

$$\frac{f}{f_1} = \frac{1}{\gamma_1},$$

where γ_1 is the magnification of the telescopic spectacles, hence

$$\gamma_1 = \frac{a_1}{a}.$$

For example, if half the retina (in diameter) is damaged, in order to restore the field of vision we have to use a telescopic system (for example, telescopic spectacles) with reduction by a factor of 2. Here the scale of space will be half what it is without spectacles; all objects will seem twice as far as they would appear to the naked eye, and the resolving power of the eye referred to the objects in question is halved - which is tolerable. But the apparent increase in distance of different objects may easily lead to an accident, particularly when crossing the street since trucks and buses will seem to be further away and their speed will appear to be less than it actually is. So an optical system of this kind cannot be recommended for use out of doors.

There is another possible way of restoring the lost field of vision known to

oculists, but (only used when the boundary between the working and damaged areas of

IT IS

the retina is roughly   a vertical line. By using the plain mirrors in

a certain position in front of the eye we can redirect the sector of the field of

vision which should have been observed by the damaged part onto the working area of

the retina. There are then two images of the left and right side of the field of

vision at the same time on the healthy area of the retina. The light reflected from

the mirrors is not so bright and slightly different in color from the light reaching

the eye directly, so that when the wearer has become used to the system, he can

distinguish between the right-hand and left-hand sides of his field of vision.

Nevertheless, the presence of fields existing simultaneously and shifting irrespective

of one another creates a particularly unpleasant sensation and a number of patients

cannot get used to wearing spectacles of this type, even after long periods of trial.

Readers may obtain an idea of the sensation experienced by a patient wearing

these spectacles if they hold a   plain mirror to one eye and look out of the

window into the street at the same time. The double load on each light-sensitive

cell in the retina officially increases the amount of information (with a highly

THE SYSTEM

noticeable amount of deterioration of quality) but is frought with a number of

unpleasant and even risky consequences (headaches, nervous disorders and so on).

However,

in discussing this problem we have gone beyond the bounds of optics and entered

the realm of neurology.

CHAPTER III

REVERSIBILITY AND IRREVERSIBILITY IN OPTICS

Sec. 1. What is "reversibility" in optics?

The reversibility or irreversibility of optical processes is of great importance for the subject interesting us - the possible and impossible in optics. The irreversibility of processes imposes a number of limitations on certain transformations of optical beams, which do not conflict with the principle of conservation of energy and are quite feasible therefore at first sight.

First of all, we must state more exactly what we understand by reversible and irreversible processes. If a process is reversible, the system in which this process occurs first in one and then in the opposite direction must end up in the initial state without the surrounding bodies experiencing any changes. This definition, which is borrowed from thermodynamics, can be used in optics as well with a fair degree of approximation.

The best illustration of a reversible process is the passage of light through a non-absorbing medium or optical system, the surfaces of which are ideally polished (there is no scatter), on condition that the reflection losses during passage through the interface are negligibly small (or, which is the same thing, that the surface is ideally low-reflecting).

Let us apply this criterion of reversibility to the said process. To do this let us place a plane, ideally reflecting mirror M in the path of the ray ABC (Fig. 21) which has passed through the interface of two non-absorbing media I and II. If the mirror M is oriented in such a way that after reflection the ray returns along the previous path CBA, nothing in the system of two media is changed and the ray emerges along the same straight line BA from which it started. But if there is refraction at point B and we make allowance for reflection on the basis of the Fresnel law, the reversibility is conditional: it exists as far as the trajectory is concerned, but not as far as the energy is concerned (when it has returned to point A the energy of the beam is less on account of losses at point B on the forward and reverse path and there are changes in the energy state of the two media).
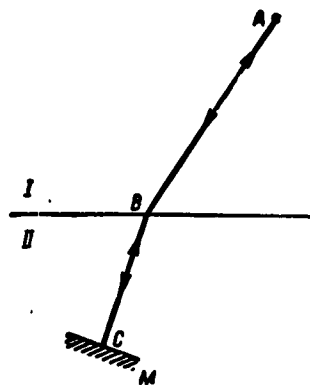
Fig. 21. Reversibility of path of ray.

Processes based on purely geometric laws (refraction and reflection) are all reversible, since we disregard losses through absorption, scatter and Fresnel reflection when we deal with them. If the path describes a definite trajectory on the forward journey, it is clear that on the return journey it describes the same trajectory. If point A is represented by a certain optical system as point A', point A' (on the return journey) is mapped by point A, and so on.

However, when we consider more general laws of propagation of light, taking into account phenomena caused by the physical basis of light rays, the problem of reversibility becomes much more complicated.

When light passes through actual and optical systems there are a whole number of resulting effects, to wit:

1) refraction through a difference in the refractive index of the medium;

2) reflection from the interface of the medium with different refractive indices (including reflection from mirror surfaces);

3) dispersion (of the prism);

4) polarization;

5) absorption during passage through an absorbing medium (light filters);

6) scatter during passage through nonhomogeneous media (dust, mist, frosted glass);

7) scatter when light falls on transparent bodies with a very fine structure (ground glass);

8) absorption at the interface between two media or when light impinges upon an opaque body;

9) diffraction due to limitation of beams, etc.

Sec. 2. Which optical phenomena are irreversible?

The criterion of reversibility as described above is not always applicable. As an illustration, let us take the case of diffraction due to limitation of the light beam. Let us assume that a beam radiated from a luminous point O (Fig. 22) is refracted by a non-aberrating lens L and forms a diffraction image O' on an ideally reflecting mirror M which sends the beam back to point O. What is the structure of this secondary image? It follows from the Huygens-Fresnel principle that at the point O there is formed a complicated diffraction pattern completely unlike the original point O; and, anyway, common sense tells us that the reconversion of the intricate diffraction pattern O' into a point is unlikely.

The answer to the question whether or not the processes are irreversible is given partly by theory and partly by experiment.

Strictly speaking, none of the optical phenomena mentioned is reversible. Refraction and reflection from an optical surface is always accompanied by absorption and scatter and it never happens that the energy of the incident flux is totally reflected or refracted. Even when the surface is thoroughly processed and coated, between 0.5 and 2% of the luminous energy is left unrefracted in the first case, and not reflected in the second; and this loss is irreversible.

The same thing happens when light is dispersed through diffraction in a prism, or when ordinary light is polarized, since some of the energy, though it may only be a small amount is scattered and absorbed.

Another irreversible phenomenon is the scattering of light, no matter what may cause it; diffraction from the edges of the lens, dust particles or drops of water in the way, a matte glass surface, or some other body.

The absorption of light is also irreversible; the luminous energy which is absorbed by the glass or some other solid heats up the latter and naturally increases its own internal energy. But the reconversion of this additional energy is out of the question.
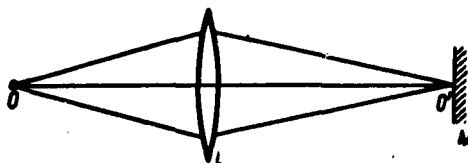
ST

Fig. 22.

Hence the passage of light through any filter, whether neutral (one which absorbs all the visible radiation to an identical extent and does not therefore change the color of the objects observed through it) or a color filter, is an irreversible process. For example, if a beam of white light becomes green after passing through a filter, it certainly cannot be reconverted into white light of the original brightness, although it is possible to convert it into white light with reduced brightness by using extra color filters.

With regard to refraction and reflection, we should point out that these effects can be considered to be reversible with fair accuracy, provided we take certain measures, such as thorough polishing and coating of the surfaces, in order to reduce losses to a minimum. The principle of the reversible behavior of rays is a particular case of these effects. It is valid insofar as we can disregard losses due to reflection, absorption and scatter.

Dispersion due to refraction in a prism is classed as a reversible effect, since the absorption by the prism can be disregarded. Newton himself proved that white light broken up with a prism can be recondensed into a slit by means of a second prism and that the white light passing through the second one is just as bright as the initial light.

These arguments apply to polarized light when there are no losses. Non-polarized light can be converted into polarized light in two mutually perpendicular planes and the latter can be reconverted into ordinary light. But here we are dealing with a rather more complex effect which will be described in detail later on.

Scatter cannot be explained by the laws of geometric optics and has to be related to physical optics. It is due to the presence of fine particles (dust, water bubbles, surface scratches, irregular surfaces forming "matte", and so on) in

the media through which the light is propagated. The smaller the particle in the path of the light beam, the more it scatters the light. According to the Huygens-Fresnel principle, the particle is turned into an independent source radiating over an obtuse solid angle: the smaller the particle, the wider the solid angle within which the light vibrations are propagated (when the size of the particle approaches the length of the wave, it acts more or less as a luminous point.

As a result of the occurrence of an enormous number of disordered secondary sources, the luminous flux, which originally issued from a virtually point source and constituted an extreme degree of   orderliness, now becomes the opposite, either gradually or suddenly (according to the orientation of the particles in the path of the ray); it turns into an infinite number of random feeble THoJewbroad rays. Although in the given case absorption losses are pretty small, the process of scatter is nevertheless irreversible; it is inconceivable that this infinitely large number of separate disorderly sources could be reunited.

As can be seen from what has been said above, the general solution of the problem of reversibility of light processes is not yet possible. The criterion given at the beginning of the chapter is not always applicable. Each individual phenomenon requires a separate approach and a special method of solution. Hence in literature we only find a small number of attempts to arrive at a general method of solving the problem. Among them a book by Jones Clark is of great interest*; the book puts forward the following idea.

Thermodynamics makes it possible for us to solve problems concerned with heat exchange. Methods have been developed for the study of thermal processes, making it possible to determine the direction in which a certain process will move. Hence, if it were possible to reduce all optical processes to heat exchange, the problem of irreversibility of any optical effect would be solved by thermodynamic methods. However, the transposition of photometric concepts of light rays, taking into account such properties as polarization, diffraction, absorption, to thermo-

---

* J. R. Clark, On Reversibility and Irreversibility in Optics. JOSA, 43, No. 2, 138 - 155, 1953.

dynamic effects itself involves great difficulties. The author has encountered a number of obstacles, in particular scatter due to diffraction, which have proved insuperable. Nevertheless, the approach he takes is worth noting.

As is well known, problems of reversibility and irreversibility are closely linked with statistics and the theory of probability. But the difficulty is that the fundamentals of statistics and the theory of probability, which come so naturally to the study of gases, cannot be applied quite as naturally to the study of luminous phenomena with their complex and varied effects.

As is well known, reversible processes are the exception, or rather an ideal *To THE ONLY* unattainable in practice. Among such processes is the passage of light through media which are absolutely transparent and non-absorbant. The *SIMULTANEOUS OCCURRENCE* of these conditions is encountered extremely rarely.

Conversely, the absorption and scattering of light are constant and universal. *PHENOMENA* The probability of these processes is infinitely greater than the probability of the propagation of light in accordance with geometric optics requiring point sources and other conditions unsatisfiable in nature.

On such a shaky foundation, however, it is difficult to formulate any serious theory of reversibility of light processes. Let us hope that the theory will be formulated and will be of use in throwing light on a number of practical problems involving reversibility, which we will now go on to discuss.

Reversibility and irreversibility of optical processes are not only of theoretical interest. Many attempts have been made to "concentrate" dispersed light radiated, let us say , by a television screen, and to direct it into a comparatively narrow beam. By doing so it would be possible to gain a great deal in brightness, since the reduction of the solid angle by a factor of $\underline{k}$ would lead to an increase in brightness of the same factor. The base of the television tube radiates within a solid angle equal to 2π . It could be reduced to $\pi/4$, which would make it possible to serve the needs of a reasonable number of televiewers, and at the same time increase the brightness by $8 - 10$ times. This is all the more tempting since an intensification of this kind is used in movie theatres by means of

special screens reflecting light solely at a useful solid angle. Why is this impossible for a television tube if it is possible for a movie screen?

Light falling on the movie screen is beamed, i.e., it is as though emitted by a projectoᴿ objective, the ᴇxɪᴛ ᴘᴜᴘɪʟ of which is small compared with the distance of the screen. Hence if we assume that the diameter of the objective is 5 cm, and the distance between the projector and the screen is 40 m, the angle of scatter of the beam of light reaching the screen is not more than 5°, i.e., the beam is extremely sharp. When they impinge upon a special screen, usually covered with little beads or tiny concave mirrors (so small that their boundaries are not visible to the audience), the rays may be beamed in any preset direction according to the position and shape of the screen ᴄᴏᴍᴘᴏɴᴇᴀᴛꜱ.

Light reflected by the matte screen surface of a television set is immediately scattered, and no optical system can gather it into a beam. A more simple proof of this impossibility ᴡᴀꜱ given ᴇᴀʀʟɪᴇʀ (Chapter II, Sec. 6).

Sec. 3. An example of the incorrect application of the reversibility principle

The reversibility principle is not always understood, even in its simplest form. Several years ago a certain inventor put forward a new method of constructing the illumination system ꜰᴏʀ a lighthouse or searchlight based on misinterpretation of the irreversibility principle. Since the example contains a lesson, we reproduce it in simplified form divested of the details obscuring the basic idea.
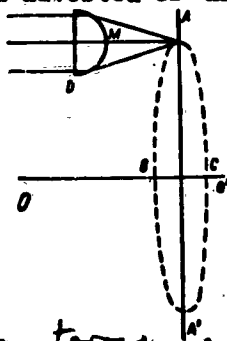

Fig. 23. Ring-shaped ᴛᴏʀᴜᴀ lens.

Let us consider an optical unit obtained by rotating the semi-circle D̲D̲'M̲ about the axis OO' perpendicular to the diameter DD' (Fig. 23). This unit is a half-torus; it is like a bread roll cut in half along the plane of symmetry (Fig. 24). If a parallel beam of light strikes this unit perpendicular to the plane formed

by the rotation of the line DD' (see Fig. 23), about the axis OO', in the plane AA'
there occurs a ring-shaped luminous line ABA'CA - which is the geometrical focal
point of the cylindrical elements DD^{V}M. The ring is particularly bright when the
source of light is the sun. Let us assume that a strongly incandescent ANNULAR
filament IS USED INSTEAD OF [the ring]. How does the beam emitted by the filament look?
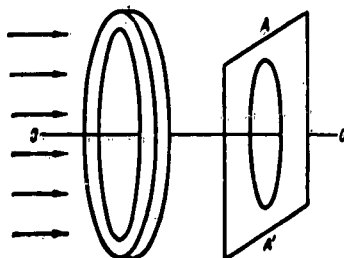

Fig. 24. Image of sun produced by ring-shaped lens.

Unduly hasty application of the reversibility principle led the inventor
to think that after refraction in the glass half-torus, the beam would become
parallel. If this were so, it would be possible to intensify the beam by adding a
number of semitoric parts forming something like a Fresnel lens; this would be very
tempting since we could THEN go on increasing the brightness of the emergent
beam to infinity. But it clearly conflicts with the Mangean-Chikolev equation given
on p. 26. Wherein lies the error? To make it easier, we will simplify the problem
still more. Let us consider the action of one element of the unit, rather than the
whole of it. More simply, let us cut the unit into two close sections containing
the axis of symmetry OO' of the half-torus. In practice, this element is no different
from a small plane-convex cylindrical lens, and so, to simplify the argument, we can
replace the complex semitoric surface by a cylindrical one. Let us assume THAT A
parallel or almost parallel beam of light (say, from the sun) falls on the plane-
convex cylindrical lens/ To make it clear, we will assume that the plane surface
of the lens is vertical. It will be possible to observe a bright line AB at a
certain distance from the lens (Fig. 25a). It is formed by the beam of rays lying
in horizontal planes. If we consider the plane vertical beams (Fig. 25b), such as
MN, M'N' and so on, after refraction through the lens L, they continue their
journey without gathering in a point. This means that the lens has no condensing
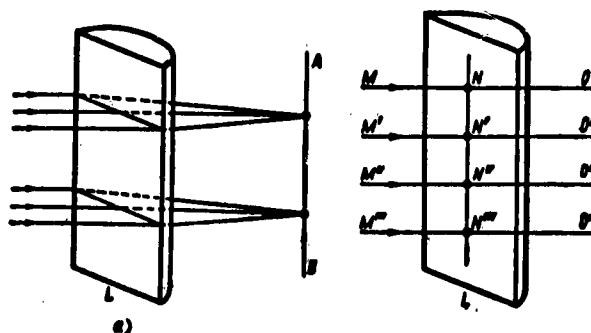power as far as the vertical rays are concerned.

Fig. 25. Behavior of rays in cylindrical lens: a) meridional; b) sagittal.

Let us now consider what happens when the beams of light are propagated in the opposite direction. By correctly applying the reversibility principle, we reach the following conclusion.

Rays proceeding from any point on the linear light source AB do not emerge from the lens as a parallel beam. Indeed, rays emerge from point M in all directions (Fig. 26). These include rays lying in planes perpendicular to the luminous filament. These leave the system in parallel beams in a direction perpendicular to the plane surface of the lens. But, apart from these,   other rays are emitted at all possible angles, such as MN_1, MN_2, MQ_1 and MQ_2. After refraction in the cylindrical lens, these rays take the directions $N_1N'_1$, $N_2N'_2$, Q_1Q_1' and Q_2Q_2', and are not parallel to each other.

Thus, the inventors expectations that the light would be concentrated in one direction were wrong because, apart from the first totality of rays forming the ring, there is another, infinitely more powerful totality. It does not take part in the formation of the ring, but merely creates a general, dimmer background which is unnoticeable compared with the bright ring. It is this fact which caused the misunderstanding.

The lesson which can be gained from this example is the following. The principle of reversibility of the behavior of rays must be applied to the totality of the rays, i .e., when analyzing any problem based on this principle, we must take into account all rays participating in the formation of the image even though they may only play a secondary role (as was the case in this example).
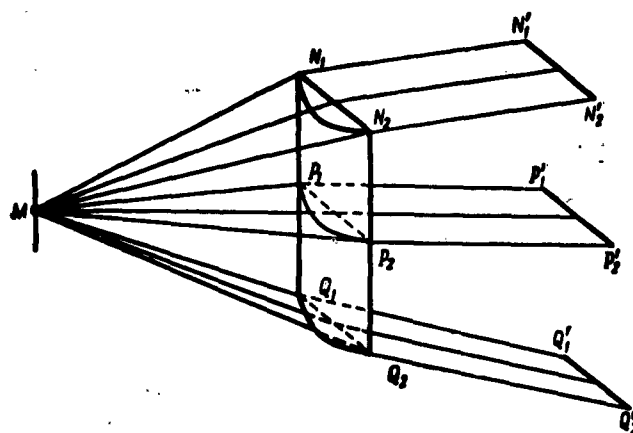
Fig. 26.   Passage of rays originating from a point
            through a cylindrical lens.

CHAPTER IV

RESOLUTION LIMIT OF OPTICAL SYSTEMS (MICROSCOPES AND TELESCOPES)

Sec. 1. General considerations

When working with present-day microscopes, biologists, doctors, metallurgists and other specialists rarely use magnification greater than 1500 - 2000. Astronomers also abide by this limit, and under normal working conditions both they and the former group of scientists *Actually use* much smaller magnification (the order of 400 - 600). It should be added that over the last half century there has been no appreciable tendency towards an increase in these figures.

Magnification is, naturally, the fundamental property of optical instruments (particularly microscopes and astronomical telescopes) used to examine and study tiny or distant objects. It therefore stands to reason that we should aim at ever greater magnification.

It is therefore no wonder that there have been so many proposals and inventions aimed at tremendous magnification by combining different optical units: astronomical objectives, microscopes, eyepieces and so on. Some of the proposals go back to the inventors school days and at least prove that the inventors learned the fundamentals of optics properly. Indeed, by means of several objectives and eyepieces we can produce a magnification of millions or billions times, and the methods for doing so are infinite, thus opening up tremendous possibilities for inventors.

It stands to reason that the astronomers and microscopists themselves have made many attempts to use high degrees of magnification, especially since they can be achieved by simply replacing an eyepiece by a strong microscope (this very simple trick produces magnification of 50 - 100 thousand times). After prolonged and laborious research to establish the most advantageous magnification, astronomers have come to the conclusion that the optimum increase in size depends on a large number of factors, among others the diameter of the objective, the s*t*ate of the atmosphere, the viewed object and, to a slight extent, the quality of the instrument itself. The *FIGURES* are small and range between 200 and 800 for the very largest

instruments. It is pointless to use greater magnification. It only leads to reduction in the field of vision, brightness and definition of the picture. ALTHOUGH /appearance THE of new details in the image is not out of the question, as we shall see later, these details bear no relation whatsoever to the object being observed.

What causes these depressing results? Let us first consider the two factors which seem the most natural - the effect of vibration of the atmosphere and reduction of the image brightness.

Vibrations of the air layers due to wind or other meteorological effects create something like mobile air lenses with a variable focus. They cause images of stars and other heavenly bodies to move about, become diffuse and blur. At greater magnifications both the vibrations and blur are increased still more and there comes a time when a further increase in magnification not only stops improving the definition of the detail, but ACTUALLY makes it worse.

The only possibility of weakening the effect of the atmosphere is to build the observatory in the hills. Allowance must be made in this case for the climate and the cloud situation. As illustrations we can point out the more important American observatories at Mt. Wilson and Mt. Palomar, built at altitudes from 1200 to 1500 m above sea level. In our country we have observatories in the ^{C}aucasus, near Abastuman and Byurakan at a height of about 2000 m.

Another factor which is sometimes used to explain the poor visibility at super-high magnification is weakness of the brightness of the image.

Indeed, whenever we look at something at night with the naked eye, in particular heavenly bodies, the diameter of the pupil of the eye Δ ranges from 3 to 8 mm. If we place a diaphragm with a circular opening with diameter Δ_1 ($\Delta_1 \leqslant \Delta$) in front of the eye, the illumination of the retina is reduced by a factor of $(\Delta_1/\Delta)^2$, which results from a corresponding decrease in the flux. The same thing happens when there is a telescopic system in front of the eye; the role of the diaphragm in this case is played by the EXIT PUPIL of the system. It follows from the properties of telescopic systems that the diameter of the pupil is γ times smaller than that

of the objective, where γ is the angular enlargement of the system.

As an example, let us take the giant refractor at the Pulkovo Observatory, which has a diameter of 80 cm and a focal distance of 14 m. Let us use a microscope with a thousand magnification as the eyepiece. This means that the focal distance will be 250/1000 = 0.25 mm. The enlargement of the entire system is 14,000/0.25 = 56,000. The diameter of the aperture is 800/56,000 = 0.015 mm, i.e., it will be 3/0.015 = 200 times smaller than the normal pupil of the eye. The luminous energy reaching the retina is 200^2 = 40,000 times smaller than if there were no optical system (disregarding the losses within the system).

What effect does this tremendous reduction in the illumination of the retina have? The number we have just quoted is staggeringly large. But we will now show that in many cases it is possible to accept this reduction in brightness.

Heavenly bodies divide up into two categories: stars AT distance so great that even after magnification thousands and tens of thousands of times they are still only points, and heavenly bodies with an appreciable angular diameter (the sun, moon, planets, comets and nebulae).

In the first case the light source can be considered a "point" (this is the only case of a point source in the application of optics). The flux striking the objective of the telescope as a parallel beam emerges from it virtually unchanged, and on condition the aperture of the telescope is less than the pupil of the eye, is received entirely by the latter. If we use D, Δ_1 and Δ to designate the diameters of the objective, telescope aperture and pupil of the eye, respectively, the apparent brightness B' of the star is related to the true brightness B as

$$\frac{B'}{B} = \left(\frac{D}{\Delta_1}\right)^2 \cdot \left(\frac{\Delta_1}{\Delta}\right)^2 = \left(\frac{D}{\Delta}\right)^2.$$

This ratio is not a function of the diameter of the exit pupil of the telescope.

Thus, the brightness of a star viewed through a telescope does not depend on the magnification, but merely on the aperture of the objective*.

_____
*This applies to a case in which $\Delta_1 < \Delta$, which is always true of very large magnification.

It is true that these arguments only hold as long as diffraction is unable to over-come the geometric effect. As soon as the diameter of the exit pupil becomes extremely small, the diffraction image of the star acquires appreciable angular dimensions and occupies a fairly large area of the observer's retina. In this case the above formula ceases to BE VALID ., and it is replaced by the law of variation in illumination in inverse proportion to the square of the magnification. Hence when viewing very bright stars, it is possible to use ultra-high magnification from the point of view of illumination (although it is pointless to do so), since there is a wide margin of brightness, but in the case of feeble stars the use becomes harmful.

In the second case, in which we observe bodies with visible angular diameter through the telescope, according to the general theory of geometric optics the brightness of the image is equal to the brightness of the object* as long as the pupil of the eye is filled with light. As soon as the magnification attains a value at which the exit pupil of the telescope is equal to the pupil of the eye, there is an reduction in the visible brightness in inverse proportionto the dia-meter of the exit pupil. Certain heavenly bodies (the sun, and even the planets and the moon) possess such a margin of brightness that a reduction of it a thousand times (and a million times in the case of the sun) still gives quite a bright image. Thus we could EMPLOY A magnification of several thousand times and still view the image with fair degree of brightness. It is only the weak bodies (comets and nebulae) which prevent us using super-large magnification even going only on: considerations of brightness of the image.

Thus, we come to the conclusion that for astronomers the weakening of the apparent brightness of heavenly bodies (at least in most cases) is no obstacle to super-magnification.

---

* This can easily be derived by comparison of the equation for a constant flux

$$\Phi = \Phi' = Bs \sin^2 u = B's'\sin^2 u',$$

and the Lagrange-Helmholtz law

$$s \sin^2 u = s' \sin^2 u',$$

from which

$$B' = B.$$

The same thing can be said of the microscopist, for at the present time he, too, has access to such powerful sources of light that he can afford a magnification of the order of tens and even hundreds of thousands of times without suffering from a defficiency in light.

We could make further attempts to explain the limitation on useful magnification by the aberrations inherent in optical systems, but it is easy to prove from the example of parabolic reflectors corrected for aberration (at least on the axis) that the explanation is also invalid.

Thus, we have exhausted all the possibilities **within the framework of geometric optics** for explaining the futility of super-magnification, and have shown that the real cause of this phenomenon, so disastrous for our attempts to understand nature, lies elsewhere.

In the middle of the last century the well-known French physicists Foucault conducted a series of experiments with optical instruments. Foucault used spyglasses with different objective diameters to view special drawings (test objects) in the form of black strips on a white background. These drawings were viewed from different distances and Foucault kept increasing the distance until the images of the strips merged into a solid gray background. At that moment the spyglass was no longer able to "resolve" the strips. The angular distance between two neighboring strips visible from the objective of the spyglass was termed the least resoluble angle.

Foucault came to the following conclusions:

1) every telescope has a certain minimum resoluble angle α ;

2) the angle α does not depend on the magnification;

3) the angle α is only related to the diameter of the objective and varies in inverse proportion to it, according to $\alpha = 120/D$, where α is expressed in seconds and D in milimeters.

Foucault carried out his experiments in the laboratory and atmospheric effects were therefore ruled out. The test objects were fairly well illuminated and the angle α was not a function of the illumination of the object within wide limits.

Thus, the factors enumerated above cannot explain Foucault's results. There must therefore be another effect which causes images to blur. This effect is the diffraction of light in optical systems.

Sec. 2. Microstructure of image produced by optical system

Geometric optics is the science of images produced by optical systems and is in fact a branch of pure mathematics. Geometric optics is based on the concept of the light ray, the Femat principle of the shortest path of a ray, and the postulate of independent propagation of light. The entire theory of images and to a considerable extent the theory of optical instruments are based on these tenets. As is well known, these theories re fairly well borne out by fact, if we keep to the more general effects and are not carried away by more exacting experiments (for example, examination of stars and microbes with super-magnification).

No matter how small the discrepancy between theory and practice, it still indicates the inaccuracy of the fundamentals of the theory. The principle of the light ray is inaccurate; the postulate of independent rays also requires more thorough analysis.

Geometric optics considers a ray as a mathematical line or infinitely thin tube carrying energy. This fiction was essential for formulating the theory; without it there would have been major mathematical difficulties which could hardly been overcome, and the theory of images would never have seen the light of day. It is better to formulate a simplified, though not entirely strict theory, and then correct the inaccuracies in the fundamentals by additions and corrections.

For what is coming we will need to rake several simple experiments. We will require for them a small piece of thin cardboard (or foil IMPERFORATED), a razor blade, a sewing needle and a strongly positive lens with a focal distance of 2 - 5 cm.

The light source requires particular attention. It must be very small (of the order of hundreds of a milimeter) and at the same time bright, since it is to be viewed by means of very fine, and therefore, very low-power rays. The best source

is the image of the sun in a highly polished ball-bearing 2 to 4 mm in diameter. As can be easily calculated, this image is from 0.005 to 0.01 mm in diameter, and its brightness is many times greater than that of any other easily accessible source (the brightness of the sun is 120,000 stilbs, an incandescent bulb filament 300 - 600 stilbs, i.e., 400 - 200 times smaller); but the sun is not at our beck and call, which makes the experiments very difficult.

The following contraption makes a good source of light (Fig. 27). The light source A (an incandescent bulb filament, or if the worst comes to the worst, a paraffin lamp flame) is projected by means of a strong condensing lens L onto a piece of thin cardboard or foil, in which a very small hole O' (preferably not more than 0.1 mm) is pierced with a sewing needle. The image of the filament A' should not be unduly reduced in size. Hence the lens L should be placed fairly close to the bulb. It is only by doing this that the hole in the cardboard O' can be filled with light. When arranging the parts, we must make certain that the image falls on the hole by covering the hole on all sides. This source can be considered a point on condition the observer is stationed **several meters** away from the hole O'.
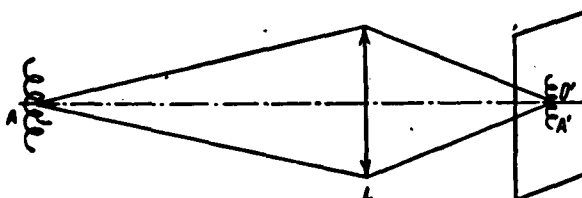
Fig. 27. Source of light for observing fraction.

A brighter picture can be obtained if we replace the point by a slit, but to do this we must have available a vacuum incandescent electric lamp with a straight filament or a paraffin lamp with a flat wick, and the plane of the wick should contain the axis of the projecting lens. The slit is cut in the cardboard or foil with a razor blade or else, still better, can be formed by two blades, one of which can move about, keeping parallel to the other (Fig. 28). This makes it possible to vary the width of the slit and gives us an idea of the effect of the width on the effect in question.

To investigate the diffraction produced by the optical system, we need *also,*

a telescopic system. Any pair of binoculars (theatre binoculars or field glasses) will do, or else a spy glass. If these instruments are not available, it is easy to make ones own telescope. For the objective we can use an ordinary spectacle lens from 2 to 4 positive diopters (for long SIGHT) and as the eye piece a strongly positive lens with a focal distance of 2 or 3 cm. It is still better to make the eye piece out of 2 lenses with focal distances of 3 or 4 cm each, placing them as shown in Fig. 29. The lenses are put in a cardboard holder and inserted in two tubes (a long one for the objective and a shorter one for the eye piece) so that it can be focused on both distant and near objects.
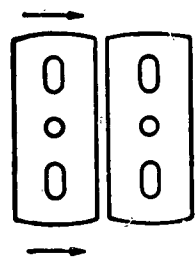
Fig. 28. Arrangement of slits for observing diffraction.
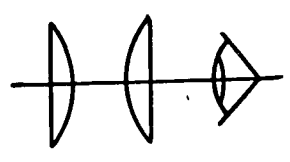
Fig. 29. Diagram of two-lens eye piece.

A telescope of this kind, provided the aperture of the objective is diaphragmed down to 2 - 3 cm in diameter, it enables us to obtain quite good images of heavenly bodies. We can observe the lunar circi, Jupiter and its four satellites, the phases of Venus, the nebulae of Andromeda, Orion and so on. Given a focal distance of 50 - 100 cm and good atmospheric conditions, we can look at Saturn and make out, if not the rings, then at least formations of indistinct shape first

noticed by Galileo through a telescope of much the same quality. As is well known, Galileo died without ever finding out what these formations were. The main reason for poor images is a phenomenon which we will now study on an experimental level; it is the phenomenon of <u>diffraction</u>.

Diffraction can be demonstrated by the following very simple experiment. If we look at a luminous point (the sun's image obtained by the above-described method in a well-polished ball-bearing*)by placing a piece of cardboard with a tiny opening (not more than 0.1 mm) in front of the eye, instead of the point we see a large white circular spot encircled by rings alternately dark and light colored. If we look at the point through holes of different sizes (the size of the hole can be judged by looking at it against the bright sky or a brightly illuminated sheet of white paper) using a strong magnifying glass, the magnitude of the spots and rings varies, a smaller opening producing larger ones. It is now easy to understand why images are not sharp at super-high magnifications, i.e., at <u>extremely small exit pupils</u>, since what we have seen is merely an image of the point in an instrument (the human eye) with a very small pupil.

Now a few words about the theory of the phenomenon observed. The fundamentals of the theory can be demonstrated by Yung's very simple experiment which can be carried out in the following way. Two thin parallel slits are cut with a razor blade in a piece of thin cardboard or foil. The width of the slits should not be more than 0.1 mm, and the distance between them should be 0.2 - 0.3 mm. A luminous point, or better, a slit (more light) is viewed through these slits, and the direction of the

* Other sources described above are considerably weaker, but given a high degree of accuracy and a lot of patience, they, too, produce good results. If the experiment does not work with the point, which may happen when the light source is very weak, it should be replaced by a slit and illuminated with the image of a lamp filament or other source, and observed through a slot not more than 0.1 - 0.3 mm in width.

two parallel slits should be parallel to the light source slit. We see closely spaced strips, alternately bright and dark parallel to the direction of the slits. The less the distance between the slits, the wider the strips appear.

The explanation, first given by Fresnel, is as follows. Each point on the source radiates vibrations propagated in all directions in the same way as a stone thrown into a pool of water causes waves on the surface (which spread in circles, the center of which is the point where the stone strikes the water. Each point which has been reached by a wave is itself set in vibratory motion and transmits the vibration to all points around it. The undular motion is propagated in this way by transference from point to point. It is important to note that each point here itself becomes a source of vibrations (the Huygens principle).
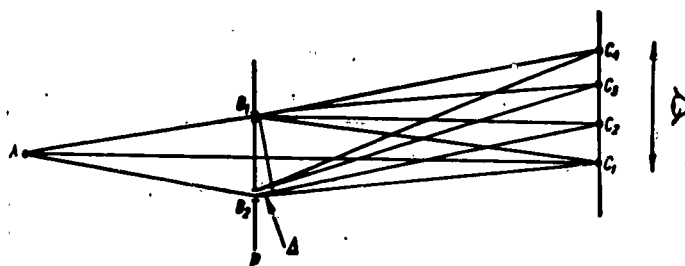
Fig. 30. Yung's experiment.

Let us further assume that using the stream P (Fig. 30) with two narrow slits B_1 and B_2, we have absorbed all the vibrations except for those passing through the slits. The moment the vibratory pulses reach them, these slits become independent sources excited (or sustained) by the same source A. Any point beyond the screen receives two vibratory impulses of equal amplitude at every moment (on condition the slits are the same width). At the points where the difference in behavior of the vibrations is equal to a whole number of wavelengths (C_1 and C_3)

$$\Delta = k\lambda \ (k-\text{ integer})$$

the impulses are added up and light is observed; at points where the difference is an odd number of half-waves,

$$\Delta = (2k+1)\frac{\lambda}{2}.$$

the impulses are subtracted and their sum is zero, i.e., there is no light (points

$\underline{C}_2$ and $\underline{C}_4$).

Simple calculations show that the BANDS observed (Fresnel bands) are equal distances apart. Their width is inversely proportional to the distance between the strips (B_1B_2) and directly proportional to the wavelength of the vibration λ and the distance between the points $\underline{C}$ and the slits. Hence, the experiments with the apertures is more effective when the slits are closer together. The source and slits should be extremely narrow or else the pattern is blurred.

It should be pointed out that this phenomenon, known as _interference_, is only observed when the two sources B_1 and B_2 emit vibrations with a time-constant phase difference, or, in other words, when they are coherent. Hence they should be excited by the same primary source, in this case the luminous point A.
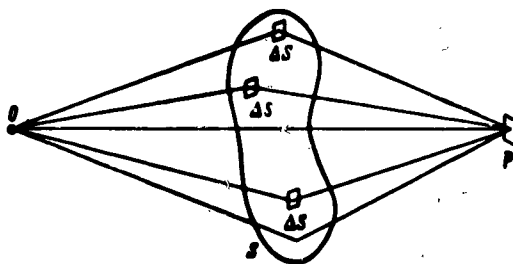
Fig. 31. Huygens-Fresnel principle.

If the screen is lit by two incoherent sources, there is no constant difference in behavior at points on the screen and the interference bands do not occur. In everyday life we encounter non-coherent sources almost exclusively. Hence interference is only observed in exceptional, artificially produced cases.

Fresnel made a bold generalisation, brilliantly borne out in practice, to supplement his explanation of interference, and indicated a method of calculating the illumination of any infinitely small area. He proceeded on the assumption that all infinitely small elements ΔS (Fig. 31) on the wave surface S send impulses characterized by amplitude and phase to the point. The vibrational state of the point P (or, more exactly, the amplitude of the vibrations, the square of which determines the illumination at point P) is calculated as the sum of all the impulses, taking both amplitude and phase into account; more briefly, according to the rule of the

addition of vectors, the absolute value of which is equal to the amplitude, and the direction is determined by the phase. A more detailed account of the Huygens-Fresnel principle (which the above-formulated rule is conventionally termed) will be given further on (Ch. V, Sec. 2).

Although the principle is simple in theory, its application involves extremely cumbersome calculations, even in the simplest cases.

Let us consider some important cases of practical interest to us. We will assume in so doing that the light source is a point located at a long distance (compared with the size of the aperture in the screen) from the entrance pupil of the optical system(or screen aperture).

1. The aperture in the screen is extremely small (fractions of a wavelength of the order of a tenth of a micron or less). The vibrations from the source reach the aperture in single phase and move on to any point $(P_1, P_2, P_3,...)$ without any appreciable phase difference on account of the very small size of the aperture (Fig. 32). Hence no where do the impulses compensate each other, and the entire space on the right hand side of the screen is flooded with light, as though the opening itself were the source; this, incidently, follows from the Huygens principle. In this case there is no image.

Exactly the same effect is produced by the extremely small point (of the same size as the hole in the screen) lit from afar from a point source. This case is particularly important and is the basis of ultra-microscopy.

2. The second, special case is important for the theory of the microscope.

Let us consider a diffraction grating, i.e., a number of transparent slits (lines) separated by opaque lines of equal thickness. The number of lines per milimeter is very great (several hundred). When a coherent beam of light impinges upon the grating, the rays are scattered in all directions and interfere with each other. Here the luminous energy is only propagated in directions for which the vibrational phases are equal or different by a whole number of periods.

Fig. 33 shows a case in which the phase difference between neighboring rays is equal to the period, while the difference [PATH is equal to one wave. The beam
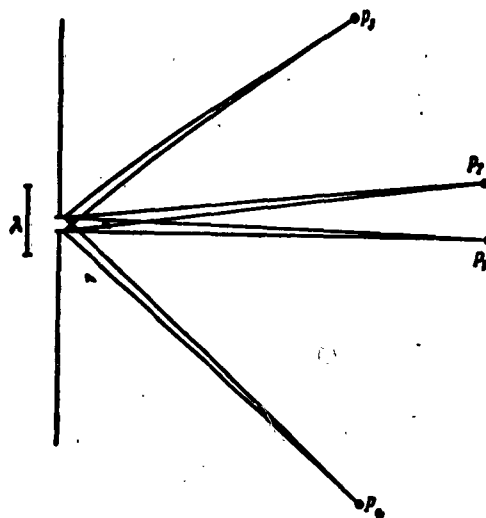
Fig. 32. Diffraction by a tiny opening.

impinging upon the grating is broken up into a spectrum. As a result several beams broken up into a spectrum emerge from the grating. The zero-order spectrum is perpendicular to the plane of the grating; the spectra of the first order formed by beams whose phase difference is equal to one period form a fan of parallel monochromatic beams on both sides of the zero; spectra of the second and third orders and so on emerge at ever larger angles.

Fig. 33. Diagram of diffraction grating.

Thus, in its effect on the parallel beam incident upon it, the grating is similar to a prism, but produces several spectra rather than just one*.

If there is an objective behind the grating, the beams gather at its focal plane in the form of plane spectra, the number of the latter being increased with the angle at which the objective is visible from the center of the grating.

3. The third case, which is the opposite of the first one and only of theoretical interest, is unrealizable; it is the case of a completely unrestricted beam emitted by a point luminous in all directions. If the point is situated in the focus of a solid elliptic mirror, an image in the form of a point without any diffraction rings is formed in the second focus of the ellipse. Unfortunately, it is not possible to have this ideal image since in order to see it, we would have to break up the mirror in some way or disturb the path of the rays in it - but then there would inevitably be diffraction.

The first case considered related to pure diffraction while/the second $CASE$ IN diffraction was totally eliminated.

Cases occurring in practice are intermediate between the two extremes. There is an image, but it is quite a different NATURE; it is by no means what we are used to from our textbooks on geometric optics.

The image in the sense in which it is understood in geometric optics should increase in sharpness as the diameter of the entrance pupil is decreased.

In actual fact diffraction causes the opposite. There is increased dispersion of the rays with reduction of the pupil; at very small apertures, the image of the point becomes a large spot surrounded by complicated diffraction patterns. The appearance of this pattern depends on the shape and size of the entrance pupil; it is effected as well by objects in the path of the rays. For example, the thin stays by which the plane Newton mirror is attached to the reflector

---

*)
As is known from the elementary theory of the diffraction grating, the number of spectra

$$k = \frac{d}{\lambda},$$

where d is the distance between the lines (constant grating) and λ is the wavelength of the diffracted light.

produce images of stars with a number of tails sticking out in different directions from the central spot. A close mesh grid placed in front of the objective appreciably scatters light and a wishy-washy image is obtained. Photographers make use of this to produce softer photographs.
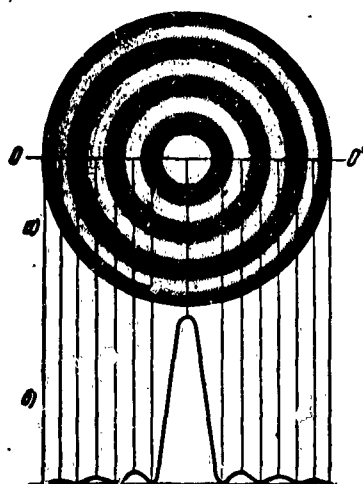
Fig. 34. Diffraction image of point: a) general view of diffraction pattern; b) curve showing distribution of illumination along line OO' through center of image.

The commonest shape used for entrance pupils in optical systems is the circle. Theory and practice show that in this case the *image of a/* (provided there is no aberration) takes the shape of a circular spot surrounded by alternating black and white rings (Fig. 34). T he radius r of the central spot can be determined by the equation

$$r = \frac{3.8}{2\pi} \cdot \frac{\lambda}{\sin u'} = 0.61 \frac{\lambda}{\sin u'},$$

(11)

where λ is the wavelength of the beam, u' is the angle at which the radius of the exit pupil is visible from the image of the point P' (Fig. 35).

If the entrance pupil is square in shape, the image of the point takes the form of a central square surrounded by other, considerably weaker squares.

A wonderful pattern is obtained when the entrance pupil takes the form of two circles A and B (Fig. 36) a short distance apart. The pattern is the same as for a single circle, except that it is streaked with dark, closely adjoining bands

and perpendicular lines joining the centers of the circles. The bands are like Young's inteference fringes, and the laws governing the distance between them are the same as those established by ^{F}resnel for Young's fringes. Here the role of the slits is played by the two circles A and B.
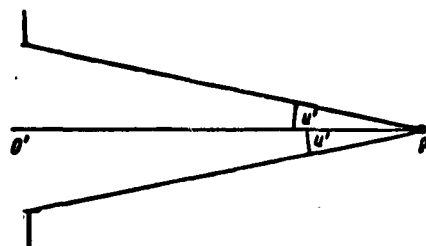
Fig. 35. Determining the aperture angle u'.
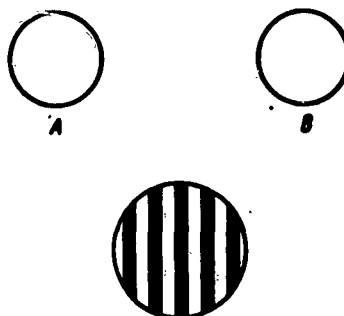
Fig. 36. Interference in a case where the entrance pupil consists of two circles A and B.

This effect can be produced by means of the home-made device described above, although, admittedly, in a different form, which does not require a particularly bright source. The source is a thin slit (0.1 - 0.3 mm) illuminated by the image of a filament or spiral in an electric bulb, or if the worst comes to the worst, a paraffin lamp wick. A diaphragm with two rectilinear slits (Fig. 37) which are such that their image fits completely into the pupil of the eye and the width of each image does not exceed 0.3 - 0.4 mm is placed in the telescope in front of the objective. The slits are parallel to the source. The telescope, particularly if magnification is large, must be kept still and correctly focused,

i.e., the images formed by the slits must coincide. It is then possible to see the image of the slit streaked with narrow interference bands in the telescope.

These examples are convincing proof of the tremendous variety of diffraction patterns obtained with different shapes of the entrance pupil. The aberrations are not taken into account here: it is assumed that there are none. This assumption, incidently, is quite legitimate for astronomical reflectors or high-power microscopes, but not for photographic lenses: the latter, which have large relative apertures and considerable angles cannot be sufficiently correct for aberration.

All these ARGUMENTS can be easily checked experimentally, using the very simple set of optical parts described above. Amazingly beautiful and varied effects can be obtained with a little accuracy and inventiveness*.
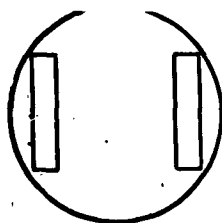
Fig. 37. Diaphragm with two rectangular slits for observing interference

Sec. 3. Resolving power of optical systems

On account of their limited size and the diffraction they cause, optical systems distort the image of a point. But a point is one element OF THE, the sum total of which make up any object viewed through an optical system. If the point is reproduced incorrectly, any object will be distorted as well.

Let us consider the image of the next simplest object after a point - two points of equal brightness, for example, two stars of the same magnitude (astronomers used the word magnitude of a star to mean its brightness)*. The two points send out incoherent light beams, i.e., ones which do not interfere with each other. Hence

---

* Those interested can find details in specialized literature, for example: R. Wood, Physical Optics, ONTI, 1936; A.A. Mykel'son, Research in Optics, 1928.

in order to illuminate a point P in the focal plane of the objective through which both stars are being viewed, we have to add the illuminations produced at this point by both beams. The illuminations, of course, are different; they are taken from Fig. 34, the position of point P with respect to the centers of both images being taken into account.
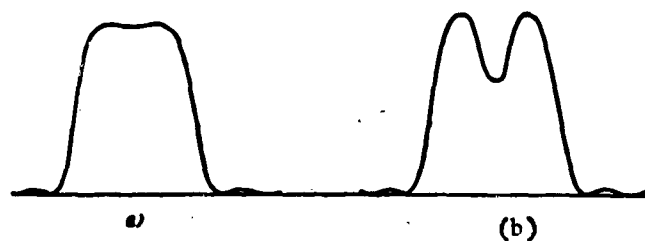
Fig. 38. Curve showing distribution of illumination at different distances between centers of images of two luminous points.

Fig. 38 shows curves for the distribution of the total illumination as a function of the position of P on a line joining the center of the images. Fig. 38a shows a case in which the distance of the centers of the images d is less than the radius of the light-colored spot r, and Fig. 38b shows a case in which d > r. In the first case the eye can only distinguish one continuous spot encircled by rings; the observer thinks that he can only see one star. In the second case, there is an appreciable dip between the two maxima. The observer can distinguish, or, more exactly, he imagines that there are two stars. It is said that the objective resolves two stars. This occurs when r ≤ d, the objective resolves and r is determined by Eq. (11)

$$r = 0.61 \frac{\lambda}{\sin u'}.$$

If D is the diameter of the objective and f its focal distance, then

$$\sin u' = \frac{D}{2f}.$$

*)

Or more exactly, the quantity describing the illumination produced by the star on the earth's surface.

On the other hand,

$$d = f\alpha,$$

where α is the angular distance between two stars in radians. Thus, we get

$$d = f\alpha \geqslant r, \quad \text{или} \quad f\alpha \geqslant \frac{1,22\lambda f}{D},$$

from which

$$\alpha \geqslant 1,22 \frac{\lambda}{D}. \tag{12}$$

Shifting from radians to seconds of an arc by a reduction of 206,000 times and by replacing λ by the wavelength of green light ($\lambda = 0.00056$ mm), we get

$$\alpha \geqslant \frac{140}{D}. \tag{13}$$

An objective with a diameter of 140 mm resolves two stars, the angular distance between which is not less than 1" (provided the two stars are of the same magnitude). Our eyes are also an optical instrument, the diameter of the aperture of which ranges between 2 mm (in bright light) and 8 mm (in darkness). The least resolvable angular distance (for two points) for a diameter of 2mm is approximately 1'. However, it cannot be greater than one minute in this case on account of the peculiar structure of the retina (there are some special cases in which it is as much as 10").

Simple calculations similar to those which we made for spy glass objectives enable us to calculate the least resolvable linear distance δ for microscopes. We get

$$\varepsilon = 0,61 \frac{\lambda}{n \sin u} = 0,61 \frac{\lambda}{A}, \tag{14}$$

where u is the aperture angle formed with the axis by the marginal ray impinging upon the microscope objective, and n is the refractive index of the medium in which the object is immersed; the quantity

$$A = n \sin u$$

is termed the numerical aperture of the microscope.

Equation (14) only relates to the Rayleigh encountered case of self-luminous objects (incandescent bodies). We usually deal with objects illuminated by a light source. The object radiates more or less coherent beams according to the type of illumination.

77

The problem of the resolving power presents tremendous difficulty in the general case and has not so far been solved, although one particular, fairly important case considered by Abbe is simple to solve. This particular case is important from the point of view of principle, and its practical interest is by no means as unimportant as it has become fashionable to think. We will therefore dwell for a moment on Abbe's theory. It only applies to objects with a thin periodic structure, for example diatoms (a kind of fine seaweed), which are of interest by themselves for the reason that they are used to assess the quality and resolving power of microscopes on account of the extremely fine nature of their scales.

For the sake of simplicity, Abbe considers as the object a diffraction grating R (Fig. 39), illuminated by a coherent source placed at the focus of the colimater L. As was described above (p. 90), the diffraction spectra S_1, S_2, S_3,...., are produced in the focal plane of the objective L_1; they play the part of secondary coherent sources. They emit vibrations which interfere with each other, as a result of which the pattern R' is formed in front of the eyepiece L_2 and is perceived as the image of the grating R.

Calculations show that the greater the number of spectra in the focal plane of the objective, the better the image is. The number of spectra is inversely proportional to the distance between the lines and the greater the numerical aperture of the microscope, the more spectra there are. There must be at least two spectra for the image structure enabling us to count the number of lines to be visible. If there is one spectrum, we only obtain a gray background, and the microscope does not resolve it. According to this theory, when the lumination is perpendicular to the plane of the grating, the least resolvable distance is

$$\varepsilon = \frac{\lambda}{A} . \tag{14a}$$

But if we apply indirect lumination, i.e., if the parallel illuminating beam impinges upon the grating, forming an angle equal to the aperture angle of the objective, the resolving power of the microscope is doubled on account of
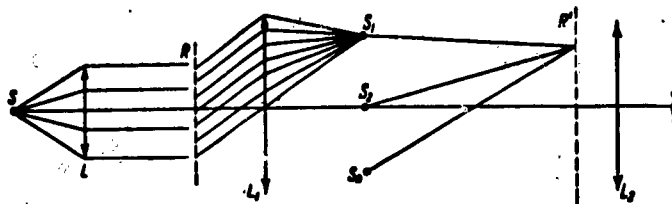
Fig. 39. Theory of resolving power of microscope for periodic
structures (point source).

doubling of the phase difference

$$s = \frac{\lambda}{2A} \cdot$$

(14b)

In this case the least resolvable distance is only slightly less than for
self-luminous bodies.

Let us observe the following fact: if the light source located at the focus
of the collimator objective possesses finite dimensions, the results of Abbe's theory
remain valid. In particular, in the case of a microscope, the light source is the
opening OO_1 (Fig. 40) of the iris located at the focus of the condenser K, which in
the given case plays the part of the collimater objective. The spectra are the
images $O'O'$, $O''O''$, etc. of this opening, produced by the objective L of the
microscope in its focal plane F'. They are circular in their form. The spectrum
formed by the rays, the phase difference between which is equal to zero, is the
brightest and has no colored fringe.

The degree to which the concept of the object image is conditional when
the instrument operates at the limit of its resolving power is shown by Abbe's
experiment for an illuminated grating, which has been reproduced by Academician
Mandel'shtan for a self-luminous grating just as successfully. If we place
several thin bands at the rear focus of the objective at the point where the
spectra are formed, we see completely different images through the eye piece, for
example, images with a double number of lines.

According to the theory of geometric optics, these bands should not have

any effect on the image, but Abbe's theory explains the distortion ACCURATELY , since the reduction in the number of spectra increases, as in Young's experiment, the number of dark and light colored fringes.
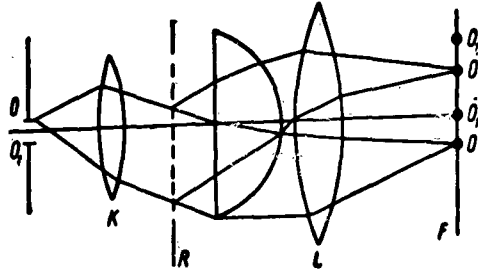
Fig. 40. Theory of resolving power microscope for structures (power of finite dimensions)

Having concluded the mathematical and rather tentative aspect of the problem of the resolving power of an optical instrument, we will now deal in greater detail with the heart of the matter.

Sec. 4. Optimal magnification of optical systems

Everything that has gone before suggests that for every optical system there is an optimal magnification (efficiency). Indeed, at excessively small magnification, the eye is not in a position to distinguish excessively small details; when the magnification is too high, we can only see a blurred pattern, further complicated by diffraction configurations which have no relation to the object being viewed. How can we select the optimum magnification in each case.

Naturally, it should be established on the basis of the properties of eyes. All the possibilities of the human eye should be exploited here. The least angular resolvable distance for the eye is 1'.

Consequently, the efficient magnification is one at which the least resolvable distance is seen at an angle of 1' after magnification.

Let us first consider some telescopic systems. Let α be the least resolvable angle of the objective. It is determined from the equation (13)

$$\alpha = \frac{140}{D}.$$

The magnification of a system through which α is visible at an angle of 60'' is

$$\gamma = \frac{60}{\alpha} = \frac{60D}{140} = \frac{D}{2.3},$$

80

and we arrive at a simple rule: the optimum magnification of a telescopic system is the magnification at which the diameter of the exit pupil is 2.3 mm.

Indeed, the diameter of the exit pupil

$$D' = \frac{D}{\gamma} \; ;$$

consequently

$$\gamma = \frac{D}{D'} \text{ и } D' = 2.3 \text{ mm}$$

Let us round off this figure to 2 mm.

At this magnification the exit pupil of the telescopic system is the same as the pupil of the human eye (in daylight). This kills two birds with one stone: we have, first, maximal magnification at which the diffraction image of the point is still perceived as a point, and, second, the greatest possible brightness of the image for all heavenly bodies (at lower magnification the brightness of stars suffers, and at greater magnification the brightness of comets, planets and other bodies visible at an angle different from zero suffers). For example, the useful magnification of the objective in the large refractor at the Pulkovo Observatory is 800/2 = 400. For binoculars with 30-mm aperture the useful magnification is 15.

In actual practice there is deviation from this rule in both directions. When it is desirable to "squeeze" everything possible out of the instrument, we have to consent to greater magnification than the rule suggests. Doubling or even (in rare cases) quadrupling the rational figures may be of benefit since it facilitates distinghishing between objects and reduces the strain on the eye. Astronomers often resort to this step which, if taken without due caution, may result in incorrect conclusions. By over-rating the possibilities of his instrument, an astronomer often tends to mistake the pattern due to diffraction and other optical illusions for reality. Many discoveries on the surface of the moon and planets were proved mere myth as soon as more powerful telescopes, i.e., ones with a greater objective diameter, began to appear.

The history of the discovery of "canals" on Mars is particularly instructive. They were first discovered by means of small telescopes, and what is especially

curious, began to disappear as observers used larger and larger objectives. When viewed through large telescopes, no channels are visible. Admittedly, we cannot assert that the canals are due entirely to diffraction: for the moment there has been no reliable explanation for them. The most likely explanation is an optical illusion by which a large number of granular formations in more or less random arrangement produce the impression of a network of straight lines when viewed from a distance. It is possible that diffraction also plays quite a part in this optical illusion, since under certain circumstances, often encountered in astronomical observation, point images are elongated into lines through diffraction. It is quite possible that by using a low-diameter telescope and employing greater magnification than the telescope can cope with, the observers stumbled on intensive diffraction effects.

Another example of an optical illusion, this time undoubtedly due to diffraction, is the volcano on Mercury, which shows up against the dark disc of the planet as a bright spot when Mercury passes in front of the sun. It has been shown experimentally that the bright spot is caused by diffraction and disappears when the telescope through which the phenomenon is viewed has a fairly large diameter.

A good example of the mistakes which are often made by people with many years experience in the field of teaching and popular science is the "ultratelescope" invented by an astronomer from Odessa. THE DESIGN of the ultratelescope, was submitted to the author of this book for his comments at a most opportune moment when he was looking for illustrations for the present chapter, and could not resist the temptation of using some of the more vivid ideas of the DESIGN.

According to the inventor's plan, the "ultratelescope" is a device which overcomes the three principal barriers, so far insuperable, which limit the unlimited magnification of optical systems. We should recall that these barriers are the decline in the brightness of the image, diffraction effects in optical instruments and vibration of the earth's atmosphere.

The inventor of the "ultratelescope" cannot be accused of ignorance: he certainly realizes the effects which he has to combat, and he knows the equations

showing their influence on their image; but he does not understand the basis of them, nor applies them correctly.

The problem of increasing brightness is solved by him with the use of a light "multiplicator" - a simple lens with a large diameter L_2 (Fig. 41) which forms at the point F_2 a secondary image of the primary image F_1 produced by the object L_1. We know that in the case of stars the increase in subjective brightness γ_{g} is

$$\gamma_B = \left(\frac{D}{\Delta}\right)^2,$$

where D is the diameter of the entrance pupil of the telescope, and Δ is the diameter of the pupil of the eye. The important thing is that this equations holds for telescopic systems of any complexity (an increase in the number of lenses merely increases the losses). The equation is a corollary of general physical laws, for example, the conservation of energy; hence none of the intermediate lenses can increase the brightness of the image, but, conversely, only reduce it on account of additional losses through reflection and absorption. The inventor of the "ultra-telescope" makes undue simplifications, taking D to be the diameter of any intermediate lens and assuming that if an intermediate lens with a diameter greater than that of the pupil is added to the objective-eyepiece system, the lens will produce an additional increase in brightness according to the equation $(D/\Delta)^2$.
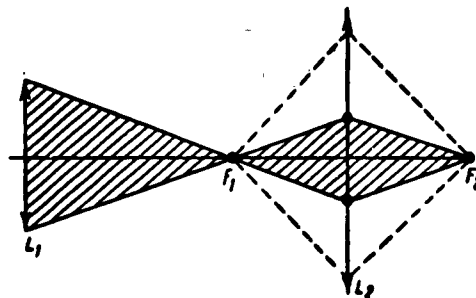
Fig. 41. Part of the "ultramicroscope" system

The SPECIOUSNESS of this is clear from general considerations. Let us endeavor to point out where the specific errors lie. There are two of them.

The inventor considers the image of a star produced by the objective to be a mathematical point, to which the equation for increase in brightness is applicable. But on account of diffraction the image of a star is a circle visible from the second

lens at a finite angle, and the formula for stars cannot be applied to it. Further-more, even if there were no diffraction and the image of the star were a point, there would not be any increase in brightness; basing his views on the Huygens principle, the inventor assumes that the image radiates as an independent source in all dir-ections and that the entire lens L_2 is flooded with luminous energy, irrespective of its diameter. But the Huygens principle only expresses optical phenomena correctly when the Fresnel addition is made, and this means that in practice the entire luminous energy is contained in a cone of rays within the aperture angle u of the objective. In Fig. 41 this cone is shaded. The rest of the lens receives such an extremely small portion of energy that it need not be taken into account. Thus, the brightness of the image does not depend at all on the diameter of the additional lens, unless the diameter of the latter is smaller than that of the light beam.

Having made such short shrift of the decrease in brightness due to very great magnification, the inventor of the "ultratelescope" does the same to the diffraction rings. It is known from the theory of diffraction that the latter is due to restriction of the light beams, i.e., to the fact that the cone has <u>finite</u> dimensions. But the inventor attributes diffraction to partial reflection of the rays from the edge of the metal frame (this misguided idea is due to the fact that certain textbooks on physical optics do not give a sufficiently clear explanation of   theory, according to which diffraction is due to a particular wave occurring on the edge of the diaphragm or frame).

The inventor of the "ultratelescope" suggests eliminating diffraction rings by means of a second diaphragm made of black cardboard which absorbs the diffracted rays coming from the edge of the metal frame. In actual fact, the cardboard diaphragm produces the same diffraction pattern as any other. In addition to that, the elimination of the diffraction rings would not in any way help to improve the resolving power of the telescope, since the former is restricted by the diameter of the central spot which cannot be turned into a point by any artful dodges*.

Let us go back to the problem of the useful, or optimum, magnification. We demonstrated above the ways in which excessive magnification could be harmful. With regard to magnification less than optimum, it can be said that it does not cause any danger. In field glasses, for example, the magnification is calculated so that the diameter of the exit pupil ranges from 3 to 6 mm, which gives good brightness when observing objects at dusk or at night, better stability of the image when the binoculars are shifted about with respect to the observer's eyes, and a number of other advantages which we will omit here since they are beyond the scope of the problems interesting us.

Sec. 5. Certain errors made by microscopists

Let us now deal with microscopes. The theory of the resolving power of a microscope is more complicated than that of a telescope, since the objects normally

* The writers comments were not appreciated by the inventor of the "ultratelescope" and his invention was put into use, though, admittedly, in a slightly unexpected way. The issue of the French journal "Astronomie" for September, 1953, published an article by the Secretary of the Astronomical Society Tequesraut, warning gullible amateur astronomers against buying the "ultratelescope" advertised for 15,000 francs. According to the inventor, the use of the instrument as an eyepiece in any objective with a diameter from 40 to 160 mm made it possible to obtain magnification from 50 to 300,000 times, with the brightness of the image and the resolving power increased into the bargain. Two samples of the instrument sent to the Secretary of the Society for his comments proved to be made with old, disused spectacle lenses, among which there were even two cylindrical ones. The name of the "inventor" was given in the article. It turned out to be a former astronomer from Odessa, who had gone abroad and (possibly not entirely consciously) was engaging there in a new form of pseudo-scientific chicanery.

viewed through a microscope are more varied from the point of view of the rays emitted by them than heavenly bodies. All objects viewed through a telescope are self-luminous (the sun, and stars), or else reflect light (planets and comets), or else do both at the same time (nebulae). But all heavenly bodies, both those which are self-luminous and those which reflect light, belong to the same category - non-coherent luminous bodies. Indeed, when we speak of two "neighboring" points on a heavenly body, we mean two points, the angular distance between which is of the nature of tenths of a second (which is at the limit of revolution of our most powerful telescopes). Such points on the sun or on planets are hundreds of kilometers apart, and in nebulae are millions, or other astronomical distances apart.

Bodies viewed through a microscope, with rare exceptions, are not self-luminous; they are illuminated by either coherent or non-coherent light, (depending on the method of illumination). But theoretical and experimental research, among which the work of the two Soviet physicists, Mendel'shtan and Rozhdestvenskiy, is of particular importance, shows that the coherence and non-coherence of illumination produces more or less the same results when determining the resolving power of a microscope. Consequently, we can apply Eq. (14) as the principal equation for deriving the least resoluble distance. On the basis of the equation, we can find the optimal magnification of the microscope.

Let F be the leading focal distance of the <u>entire</u> microscope. As is easy to see, it is equal to

$$F = \frac{f}{\beta},$$

where f is the focal distance of the eyepiece, and β is the linear enlargement of the microscope objective. Let u be the aperture angle of the microscope. According to the definition of focal distance, the following relationship exists between the angle u, the radius of the exit pupil $D'/2$ and F:

$$\sin u = \frac{D'}{2F}. \tag{15}$$

Just as for telescopic systems, the useful magnification is considered the magnification at which the diameter of the exit pupil D' is equal to 2.3 mm, since this value corresponds to the least angle resoluble by the eye - 1'. Consequently,

$$F = \frac{1,15}{\sin u}.$$

On the other hand, the magnification of the microscope is conditionally 250/F', where F' is the rear focal distance of the microscope, and F' = F/n, where n is the immersion refractive index. Hence the useful magnification of the microscope Γ_0 is

$$\Gamma_0 = \frac{250n}{F} = \frac{250n \sin u}{1,15} = 220n \sin u = 220A. \qquad (16)$$

Thus, the optimal enlargement of a microscope is 220 times its numerical aperture. Since the latter is never greater than 1.4 - 1.5, in the most powerful immersion apochromates, we must conclude that the useful magnification of a microscope is not more than 300 - 350. Here, too, as in telescopic systems, we can double or even treble these figures. Nevertheless, magnifications greater than 1000 are clearly pointless and even parmful: they result in obvious diffraction which adds its own pattern to the outline of the object being viewed and causes all kinds of errors and misinterpretations.

Generally speaking, inadequate knowledge of optics may cause not only young, inexperienced scientific workers, but also scientists of world standing to make mistakes, sometimes very serious ones. A number of objects of tremendous interest in biology, zoology, cytology have dimensions slightly smaller than the least resoluble distance. If the microscope is handled skillfully, the objects can be detected, but, obviously, it is extremely easy to fall victim to optical illusion in the process. There have been frequent instances of this, and they will be repeated until all persons working with microscopes realize that viewing images through the eye piece of a microscope without understanding the theory of it is just as difficult as reading a book in a language of which they only have a smattering.

Let us look at some examples of errors in interpretation in optics, which Professor Ya. Ye. Ellengorn has kindly communicated to me and which are of great interest.

As pointed out above, oblique illumination produced when the diaphragm of a condenser is shifted sideways may double the resolving power, which, naturally

enables us to obtain greater detail in the image, particularly when the object has a periodic structure. During observation, however, we must show caution and not jump to conclusions. Under no circumstances must we do what is recommended by Belling in his book "Working with a Microscope",(New York, 1931), in which, among useful advice, he insists on avoiding oblique illumination and that, for greater accuracy, the diaphragm should be kept absolutely still by means of wire. This precaution, however, is not a guarantee against certain kinds of optical illusion.

A similar type of error is the suggestion made by a Soviet author (in the journal "Plant Growing" for 1952) that the focal plane of the objective should be covered with a diaphragm, leaving only a third of the area of the exit pupil visible so that the definition of the photographs obtained is improved. This device reduces the aperture of the microscope and therefore reduces the resolving power of the objective, no matter what the structure of the object, and the number of distinguishable details is decreased. The author probably used an objective with high residual aberration and the images became "sharper" after the diaphragming operation, i.e., they were freed from background and halos, thus creating a false impression of definition. This method is permissable only when the object being viewed has a coarse structure.
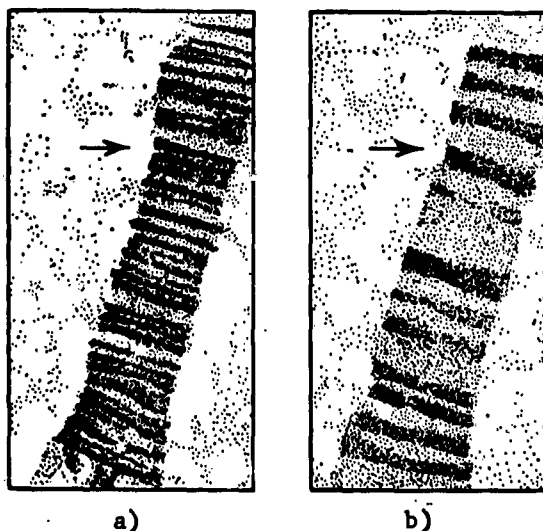

a)   b)

Fig. 12. Section of chromosomes from saliva glands of the fly Drosophila Melanogaster photographed with the same objective with different apertures.   88

As an illustration, we show two microphotographs (Fig. 42) of the same area of chromosomes from the salivary glands of the fly Drosophila melanogaster, taken by Ya. Ye. Ellengorn. Both photographs were taken with an objective with a numerical aperture of 1.3 and linear enlargement of 90; the total magnification is 3000. The photograph 42a was taken in the normal, i.e., the focal plane of the objective was entirely filled with light while 42b was taken with the same optics, but a diaphragm with a narrow aperture was placed at the level of the focal plane of the objective, producing a considerable reduction in the numerical aperture (approximately to 0.2). This resulted in certain fibers, clearly resolved and easily visible in the left-hand photograph, merging together and becoming indistinguishable in the right-hand photograph. The area marked with an arrow in both photographs is particularly characteristic.

The way in which even great experts over rate the possibilities of the microscope is shown in an article by a prominent Soviet botanist on the morphology of chromosomes.

Having developed a method of measuring chromosomes by making sketches using an Abbe apparatus (an apparatus which projects the image onto a sheet of paper, after which the outline of the object is drawn with a pencil), the botonist gives the magnitude of the chromosomes with an accuracy of one hundredth of a micron.

He bases his idea on statistical methods which make it possible to assess the most probable dimensions in question with great accuracy on the basis of a large number of observations. However, no matter how extensive the material which the writer uses, the accuracy which he banks on is clearly exaggerated, for even with the best possible microscope we cannot determine the length of an object more exactly than a third of a wave, i.e., 0.2 microns. It should be taken into account that when making sketches this degree of accuracy cannot but suffer, since in drawing the actual outline there is an element of arbitrariness, i.e., a systematic error which not even the use of statistical methods can eliminate. Hence it would be more cautious to assume that the size of the chromosome, even on the basis of a great deal of statistical material, cannot be determined within greater accuracy than

0.1 - 0.2 microns.

A comparison of these two publications, which relate to different fields, though they are linked by the same guiding idea, is extremely instructive. It is odd that in the first one, the writer, who used a perfectly correct method, was later ridiculed and criticized by the authorities, while the unjustified method in the second publication did not apparently give rise to any objections.

The first article was published back in 1869 in Germany by the young botanist Flegel in the "Botanical Journal" and testifies to his great understanding of the formation of images; it is all the more surprising that at that time the work of Abbe and others had not yet been published. Studying drawings of several types of diatoms with periodic structure, the author obtained a great deal of valuable data on the structure, in particular its period, without even resorting to the microscope, but by examining the spectra caused by this structure through a spy glass. He obtained such perfection that his results were confirmed in later years, and with electronic microscopes at that.

Flegel employed the system of comparing diffraction patterns produced by the structure he was studying and the pattern of the spectra produced by larger-scale structures directly visible through the microscope. Structures similar among themselves but possessing different periods produce identical spectral patterns, differing only in angular dimensions. This is a perfectly legitimate method and can provide much more than observation through the most powerful microscopes; this has been confirmed by later research.

The second writer, Nevel, in an article on chromosome structure published in 1952, describes a method which at first sight seems very similar to Flegels. The structures observed by him (slightly tinted fibers with a thickness at the limit of the resolving power) are somehow interwoven. But how? In order to find out, Nevel made fibers out of glass, interwove them in different ways, photographed them with an ordinary camera and compared them with microphotographs obtained with microscopes, paying heed to the distribution of light and shadow.

This method is completely wrong. Chromosome fibers produce diffraction

effects in a microscope which completely distort the photographs. But the photographs of glass *TUBES* which are so large that they do not produce any appreciable diffraction pattern are not the least like the microphotographs. Neyel's method is wrong, although from the point of view of "common sense" it seems more valid than the indirect, but quite correct method used by Flegel.

The case of diatoms of the genus nitzschia is very curious. This genus divides up into species, and the pattern on the outside surface of the shell of these diatoms is usually used as the distinguishing feature. The arrangement and shape of the lines in the pattern are used as a basis for attributing the observed specimen to a particular species. The structure of the pattern is extremely fine and lies on the border of the resolving power of the most powerful telescopes, hence microscopic technique here acquires enormous importance.

In his work* Ya. Ye. Ellengorn has shown that in actual fact all Nitzschia have the same structure and that all the varieties of structure observed by different people are in actual non-existent.

Let us see what the author himself says about this.

"When determining diatomic seaweed under the microscope, the structure of their shells made of silica is of decisive importance. As an illustration of the curiously conditional nature of the microscopic images which are usually shown for the structure of these objects, let us look at the following facts.

The genus Nitzschia is characterized most frequently by very long, narrow shells without any central seam, but with special "Kielpunkts". Furthermore, the surface of the shell (most probably inside) is covered with a very fine structure. The diffraction spectra of all Nitzschia obtained in the focal plane of the microscope objective are absolutely the same type in principle. Apart from a zero maximum, the spectra, when the lighting conditions are correct, show four more diffraction maxima of the first order, arranged in each QUADRANT of the circumference. This diffraction structure suggests that, generally speaking, the elements of the fine structure engendering them are located at the corners of the square. And indeed, some of the seaweeds are known to have a fine structure in the form of "lines" RUNNING up

and down the shell and intersecting at right angles.

But in a whole number of cases the Nitzschia shells form different images: they show a system of lines running across the shell, and no longitudinal lines at all; hence, instead of the "network" there is only transverse shading. In other cases this transverse "shading" appears in such a way that the "line" seems to consist of a number of "points" and is not continuous. Finally, in the case of a number of shells it is not possible to detect the structure at all. All these structures, obviously, are important in systematizing these objects.

Nevertheless, they are all based on misinterpretation. Any of the above-described structures can be obtained for any Nitzschia shell and depends on the degree to which the diffraction maxima of the first order are included in the focal plane. If it has not been possible to include them in the focal plane at all, there will be no structure on the shell; but if only two of the four have been included, either transverse or longitudinal lines can be obtained as desired, and by introducing all four maxima, we get either transverse "shading" except that the "line" consists of a number of points, or else a system of intersecting longidutinal and transverse lines.

It can be proved, however, by certain experiments that in actual fact the shell has very fine structural elements of an oval shape situated more or less at the corners of the square, and that the longitudinal direction of the ovals is more or less perpendicular to the length of the shell. In other words, the structure closest to the truth is the one in which the transverse lines divide up into individual points".

These views may be proved from another example. Let us consider Fig. 43. The bottom of this drawing shows four microphotographs of the same diatomic

Ya. Ye. Ellengorn. Experience in studying the fine structure of certain diatomic shells. Botanical Journal, 1940.

preparation Frusturia rhomboides var. saxonica, taken by Ellengorn. All the photographs show the same shell when viewed through a microscope with an objective 1.25 x 90, although the methods of illumination are different. At the top of Fig. 43 we see a diagramatic representation of the focal plane of the objective for different positions of the condenser diaphragm.

In the first one (43a) the shell only shows a general outline and a vague central seam; there is no fine structure. In the focal plane we only observe a zero maximum 0, in other words, the image of the shell occurs when there is narrow central illumination and is due solely to the light which is radiated by the source of illumination.

In the second photograph (43b) the shell is streaked with a number of bands running parallel all the way down. This structure is found when the zero maximum is shifted perpendicular to the length of the shell over to the edge of the focal plane, which can easily be done by means of Abbe's illuminating device. Here we have to add a diffraction maximum of the first order p to the focal plane. If the zero maximum is shifted from right to left along the diameter of the focal plane until it reaches the area of the maximum p, a first order diffraction maximum p' would appear instead of the zero maximum.

In both cases longitudinal strips are found on the shell. The addition of the diffraction maximum of the first order makes it possible to obtain an image by means of the light emitted by the illumination source (zero maximum) and the light which occurs on the actual structure due to diffraction.

In the third photograph (43c) the lines run across the shell. This structure occurs when the oblique illumination is beamed in such a way that the zero maximum in the focal plane shifts parallel to the length of the shell. In this position it is possible to introduce diffraction maximum of the first order into the focal plane and the maximum is designated n to distinguish it from the previous letter. When the position of the zero maximum is diametrically opposite, it is possible to add the first order maximum n' to the focal plane, and obviously this time we find an image of the shell with transverse striation.

Finally, the fourth photograph (Fig. 43d) shows the shell with a grid-like structure; here the two structures which appeared separately in the earlier conditions are now combined. In this case the position of the zero maximum in the focal plane is such that both first order diffraction maxima can be added to it. On the pattern produced by the focal plane it can be seen that the maxima p and n are not part of it when the illumination is in this direction.

The example quoted by Ellengorn is so convincing that it needs no further comment.
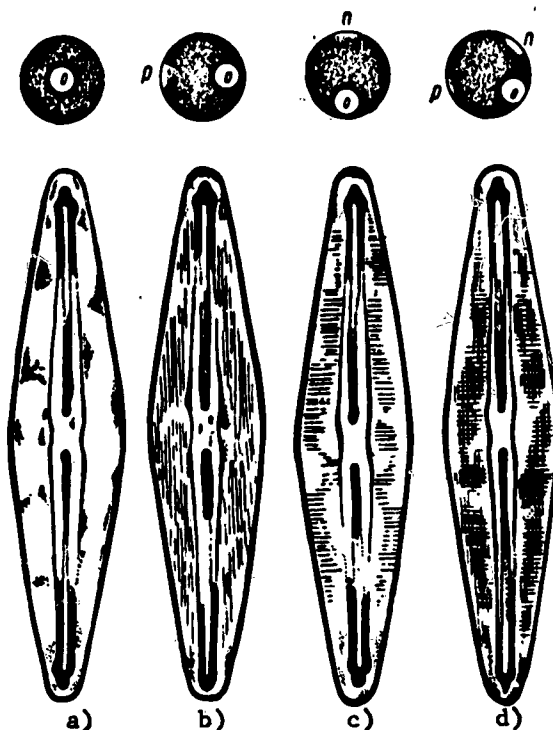
Fig. 43. Armor of the diatom Frusturia Rhomboides var.
Saxonica, observed at different positions of the
condenser diaphragm.

Sec. 6. What can be seen by means of contemporary optical instruments?

Let us first determine the limit of the possibilities of optical instruments. Let us see what the optical instruments which we possess at the present time can show us, and let us analyze the prospects for using them to unravel the secrets of nature in the future.

94

As before, let us first consider telescopic systems from the standpoint of observing celestial bodies, and then let us deal with microscopes.

For telescopic systems, the least resoluble angle in terms of radians is determined by Eq. (12)

$$a = 1.22\frac{\lambda}{D} .$$

from which it follows that to reduce the least resoluble angle the only thing we can do is to increase the diameter of the objective or reduce the wavelength of the rays producing the image.

Indeed, the diameter of objectives is increasing as the machinery used to produce them is being improved. The largest of the objectives used at the present time is part of the reflector at Mt. Palomar Observatory (USA). Its diameter is 5000 mm and enables astronomers to resolve 0.05" (50 m on the surface of the moon, 8 km on the surface of mars when closest to the earth). This does not mean that it cannot be used to view objects the dimensions of which in the transverse direction are less than those mentioned (we will explain later what dimensions of luminous bodies are visible). It only means that the image of an object of this size is no longer a shapeless spot - we can sense its shape (though only very roughly) and note whether or not it is elongated, although we cannot distinguish a circle from a square unless their dimensions exceed those given by several times. The missile "Columbia" which orbited the moon in Jules Verne's novel "Around the Moon" could not have been visible through the enormous telescope at Mt. Wilson Observatory, but it could be observed as a point against the black background of the moon if the sun were in a favorable position with respect to the missile.

Let us deal with sources of light which can be observed, but only in the form of dots. We shall see that in a case in which the problem of the shape does not arise, the possibilities of optical instruments are tremendous. Let us calculate the distance at which a [SIGNAL FROM A searchlight with a 2 m diameter illuminated by a high intensity arc (brightness 100,000 stilbs) would be visible. According to the Mangean-Chikolev law, the light intensity of a searchlight of this kind is equal to 100,000π x 100^2 candles = 3 x 10^9 candles.

Let us apply the following facts from photometric textbooks. A candle is visible with the naked eye at a distance of 27 km*). Consequently, 3×10^9 candles are visible at a distance $\sqrt{3 \cdot 10^9}$ greater. Furthermore, if we use a large hundred-inch reflector to observe the signals, the useful magnification of which is $2500/8 \approx 310$**, our object is visible from a distance greater by a factor of 310. Consequently, it can be seen from $27\sqrt{3 \cdot 10^9 \cdot 310}$ km = $5 \cdot 10^8$ km. This is three times the distance between the sun and the earth; it is almost equal to the distance between the earth and Jupiter. The observation of signals from the earth on Mars is therefore quite a simple task. Unfortunately, the Martians cannot return the compliment, since their planet always has its illuminated side facing us, and they could hardly make out a weak signal against a bright background. If a 5 m mirror telescope is used, the distances given above can be doubled.

In order to discover a "satellite" 50 cm in diameter revolving around the earth at a distance of several hundred kilometers, all we need are field glasses and favorable illumination.

Signals from the moon are possible by means of small projectors. But two lights at a distance of 50 m apart would look like one light with double brightness.

Let us now deal with microscopes. The equation determining the least resoluble distance ε in the case of self-luminous bodies takes the form /see equation (14)/

$$\varepsilon = \frac{0,61\lambda}{n \sin u} \, .$$

(17)

No matter what system of illumination is used, it is only the numerical coefficient which varies in Eq. (17), and even then only slightly.* Assuming $\lambda = 0.00056$ mm, and $\underline{n} = 1.5$, we obtain 0.0005 mm for ε, i.e., using very strong immersion microscopes we can distinguish two objects half a micron apart.

*)
Naturally, on condition that the darkness is total and the eye has had a long time to adapt.

**)
According to the equation $\gamma = D/D'$, D' being taken as 8, since we are concerned with the diameter of the pupil of the eye in total darkness.

96

If we only want to find an object without concerning ourselves with its shape, the size of the object can be hundreds of times smaller. Using a particular device, the ultramicroscope (Fig. 44) which only differs from the conventional kind in the method of illumination --we can view particles of the order of thousandths of a micron. The particles are strongly illuminated by the lateral beams P and P' which do not impinge upon the objective O, and do not therefore create a bright background. Light scattered by particles A strikes the observer's eye and he sees an image against a black background, like stars in the sky.

Ultramicroscopes enable us to see particles the size of which does not exceed 5 milimicrons.

Equation (17) shows that the least resoluble distance is proportional to the wavelength λ. Thus we can reduce δ by using ultra-violet light. Admittedly, for this purpose we cannot use the conventional glass objective, since it does not transmit ultra-violet light.

We can use special objectives made of transparent crystals, (quartz or fluorite) or, still better, mirror systems. Since our eyes are not sensitive to ultra-violet rays, the image has to be photographed, and this is the great difficulty with these microscopes. On the other hand, their resolving power is approximately double what it is in conventional microscopes, for the same numerical aperture.
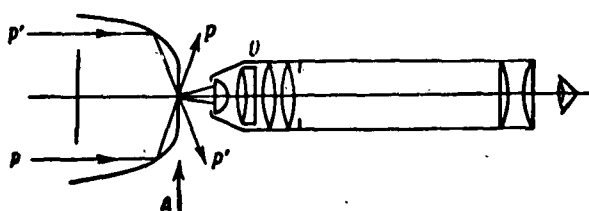

Fig. 44. Observation when the field is dark

Sec. 7. Future prospects

What are the prospects at the moment for improving optical instruments? The lease resoluble angle α in astronomical instruments (reflectors and

refractors) is determined from the equation we have frequently used, i.e.,

$$a = \frac{1.22\lambda}{D}.$$

This angle can only be reduced by increasing the diameter of the objective D and reducing the wavelength λ. Furthermore, we must eliminate the troublesome vibrations of the atmosphere. There are two ways to meet these requirements: 1) increase the diameter of the objective, and 2) build observatories high in the mountains, which reduces the atmospheric vibration and absorption of ultra-violet rays at the same time. The astronomical telescopes of the distant future will probably take the form of reflectors (aluminum-coated mirrors do not absorb ultra-violet light) tens and hundreds of meters in diameter, revolving on tires 20 or 30 km high, containing the observers and laboratory.

Now that artificial earth satellites are being built, it is conceivable that a satellite of fairly large size could carry, if not a complete observatory, at least one powerful instrument, either operating automatically or else controlled by an observer. Under such circumstances the harmful effect of the atmosphere would be eliminated and the instrument would, furthermore, be able to operate night and day without stopping since the scattering of the sun's rays by the atmosphere would be eliminated too.

Furthermore, there are still other possibilities for counteracting the effect of the atmosphere - automatic compensation of the turbulence. In this age of automation such possibilities no longer lie within the realm fantasy.

Further improvement in microscopes is possible if we use materials and fluids with high refractive indices. For the first time it is becoming possible to use diamond (the refractive index of which is 2.4), which steps up the resolving power of the microscope by a factor of 1.5. Another quantity contained in the denominator of the equation for ℓ, $\ell = \frac{0.61\lambda}{n \sin u}$, namely sin $\underline{u}$ has already attained virtually its maximum -- unity.

Reduction in wavelength also means improvement in the quality of microscopes. But this possibility should not be abused SINCE immersion fluids cease transmitting rays in the distant, ultra-violet region of the spectrum. A great increase

in visibility can be expected from the devolopment of efficient systems of illuminating the objects being viewed. This problem has been little studied so far.

It can be concluded that it would be unreasonable to hope for considerable advancement within the next few years, or even next few decades. In many respects (aperture, wavelength) the ceiling has been reached and further improvement can only be achieved in the smaller details: improving methods of illumination, improving the quality of the image, and so on.

But no matter how depressing these conclusions may seem, they do not mean certainly that if we remain within the bounds of optical instruments (and do not resort to electronic instruments) we will be unable to delve more deeply into the realm of tiny, distant objects. The possibilities of optics are still not exhausted.

Sec. 8. New ways of using optical systems

The verb "to see" is so commonplace that when using it, no one thinks very much about its exact meaning. For many people "seeing" means receiving an image of certain objects on the retina. This is why there is frequent surprise at the fact that objects do not appear upside down.

In actual fact, the image on the retina is only the beginning of "seeing". The principal work of processing and understanding the reactions which occur in the retina takes place in the nervous system and the brain. If we were able to obtain an enlarged photograph of the pattern created on the retina when we look at something, we would be amazed at the low quality of it. We would be struck particularly by the great lack of contrast, even appreciable at the center, in the region of the yellow spot, and which becomes extreme in the immediate neighborhood of the spot. This is the result of aberration, the existence of which in the eye can be shown by special experiments. We would see a blank spot (blind spot) and a number of other defects in the photograph.

But we do not see anything of this kind when we look at an object. Furthermore, if our eyes are normal, or we wear carefully selected glasses, our vision always produces an impression of ideal definition; we cannot imagine anything better.

99

As regards the dark spot, even if we know about its existence, we cannot locate it no matter how hard we try (unless we resort to the special experiments like Mariotte's with a cross and a circle, described in all textbooks on physics).

It follows from this that the brain excludes from the image on the retina everything that does not relate to the pattern and depends on defects in the eye. This cutting out effect is not a conscious one, of course.

It is possible to cite other examples showing how the brain processes impressions produced on the retina. Fig. 45 shows a cube, cylinder and sphere. In actual fact we are dealing with flat drawings. But the arrangement of the shadows and the perspective suggests - with the assistance of our consciousness - that the figures have volume. The perception of relief by an eye is also the result of consciousness which makes allowance for the distribution of shadows, angular dimensions and so on.
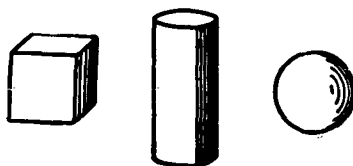
Fig. 45. Perception of relief due to shading.

We can go still further in this direction. Let us look at a photograph taken from an aircraft of territory on which manoeuvres or military operations are taking place. At points where the eye of an inexperienced observer sees nothing but a shapeless spot, the reconnaisance scout finds a tank;;some scrub, completely EMPTY /for the former, is found to be full of soldiers by the latter, and so on. Copious experience and constant work with photographs have improved the "acuity" of the observer's vision many times. Here, too, there is unconsciousness or semiconsciousness by the brain as in the first two examples, plus the use asidiously acquired knowledge. These are exactly the circumstances under which the astronomer works with the telescope or the biologist with a microscope. The image which he examines

comprises far more than they can see directly with the eye. But one has to know how to see "far more". And to do so we need a thorough knowledge of the theory of images.

Let us begin with astronomers. The theory of images of heavenly bodies is simple, but unfortunately, the entire simplicity is reduced to nil on account of the atmosphere. Let us disregard this defect for a moment and also the aberrations in the optical instruments. The image of a star located at infinity should under these circumstances appear as a bright spot surrounded by light and dark rings. The distribution of the illumination over the total area of the image is known exactly (see Fig. 34).

The slightest change in the object, for example, an other star, although its distance may be many times less than the diameter of the mean image (i.e., closer than the least resoluble distance) means a change in the image pattern. Naturally, this change is so slight that it cannot be detected by the eye (although it can be assumed that the expert who has mastered his trade could "see" the second star). But by means of sensitive photometer sounds measuring the illumination of each point of the image, the change can be observed and deciphered, i.e., laborious calculation determines the position of the second star and its illumination. The same thing applies to a star with an appreciable width, i.e., hundreds or tens of times greater than the diameter of the central diffraction spot. This width does not cause any appreciable change in the pattern observed by the eye, but the distribution of illumination over the spot and over the rings is altered.

Any deviation from a point - the presence of other, close-lying stars or satellites, the appreciable angular dimension of a star, irregular shape, the presence of rings as in the case of Saturn, or deviation from spherical shape, as in the smaller planets, - has an effect on the distribution of illumination. It has to be admitted that this way of measuring the variation in illumination is for the moment a dream. First of all, the effect of the atmosphere and aberration in the objective blur somewhat the diffraction pattern, as the French physicist, Boisse rightly points out in his book "Diffraction": "Many astronomers assure us that they

have never seen these notorious rings through their telescopes, although they are easily observed in the laboratory".

The second obstacle in the way is the tremendous difficulty involved in measuring the illumination distribution. The photometer-sounds required are only just beginning to appear, their sensitivity is low and their power to determine the illumination of very small areas are far from adequate.

A third obstacle is the difficulty of deciphering the data, i.e., determining the illumination distribution from the given pattern. Furthermore, there is no certainty of solving the problem uniquely; in other words, there may be several configurations corresponding to the same distribution.

Since we are devoting particular attention in this book both to the frequently encountered and also theoretically possible errors, we should point out one more which may easily arise when seeking ways of using the diffraction pattern most effectively, i.e., increasing the pattern to a considerable size so as to be able to study it more easily. To do this it might seem that we only need examine the diffraction pattern through a microscope, but this will lead us into a classical error - the use of beams with small apertures to obtain large magnifications.

The basis of the error can be explained in the following way. The diffraction pattern which we intend to study under the microscope emits rays contained within an extremely small solid angle determined by the aperture angle of the objective of the telescope. In large refractors this angle is of the order of 1/30 (2°); in a reflector it will be more, but usually not more than 1/12 - 1/20. In other words, the diffraction image of a star viewed as an object radiates a beam of rays into the microscope with a small aperture angle; the numerical aperture $A = n \sin \underline{u}$ of this beam is not more than 0.10 - 0.08. But according to the rule laid down for microscopes the useful magnification is 220 A, which in our case is 20 - 22. This is such a weak magnification that the microscope can be replaced with a simple eye piece with a focal distance of 12 - 11 mm, in which the diffraction pattern is just about visible. If the magnification is greater, the diffraction in the eye of the observer comes into play and the image of the pattern produced by the objectives is

102

distorted.

The futility of large magnifications can be further explained in the following way. The useful magnification of a microscope (and also a telescope) is such that the diameter of the exit pupil of the optical system is 2 mm. When the magnification exceeds this limit, the diameter of the exit pupil drops below 2 mm and begins to "diaphragm" the eye. As long as the diameter of the pupil is greater than 1 mm, there is no particular trouble, except for a drop in the illumination of the retina (hence some writers point out that the useful magnification of a microscope is 400 - 500 A). But if the pupil is further decreases, diffraction and other effects in the eye begin to play a considerable part. On account of the narrowness of the rays incident on the lens of the eye, the inhomogeneities of the latter appear, so to say, as sharply defined spots on the retina; even dust particles floating past the cornea form shadows and the pattern of the shadow spots, and so forth, is added to the diffraction effect on the retina.

The same results are obtained when taking photographs with a microobjective since the latter produces a diffuse, blurred image on account of the extremely small relative aperture. The solution to the problem of improving the resolving power has to be sought along other lines.

As an example of the correct way to tackle this problem we can quote the instrument in which diffraction is used for the purpose of improved "vision" of the object, particularly stars, the angular diameter of which is not vanishingly small. The instrument was put forward by the American physicist Michaelson and the principle of it is shown in Fig. 46.

To mirrors M_1 and M_2 at some distance from each other (up to 10 m) are attached to a telescope. The latter receives beams from the star being studied for angular diameter. The focal point of the objective OO_1 receives a picture similar to that shown in Fig. 36. If the star has an angular dimension visible through the telescope with objective with a diameter equal to the distance between the mirrors M_1 and M_2, the light bands in the interference-diffraction pattern widen,

103

th.... ...nds become lit and disappear at a certain relationship between the distance M_1 and M_2 and the diameter of the star. This device has made it possible to measure the diameters of a large number of stars.
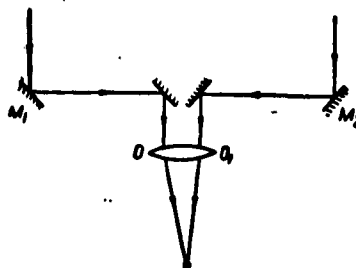

Fig. 46. Michelson's arrangement for determining the angular diameter of stars.

A disadvantage of Michelson's method is difficulty in deciphering the data: a double star gives results similar to those obtained for a star of finite width. In order to find out which of the two possibilities is the correct one, the mirrors have to be rotated about the axis of the telescope for further observation. If the star is a double one, the diffraction pattern changes in appearance during the rotation; for example, when the mirrors are perpendicular to a line joining both stars, the same pattern is obtained as for a star without an observable diameter.

Michelson's system can be made still more complex by increasing a number of mirrors to four or six and placing them at the corners of regular polyhedra. Then, the interference-diffraction pattern becomes more complex as the area of the mirrors is increased, and approaches more and more closely the pattern which a solid mirror with a diameter equal to the distance between the outside mirrors would produce. But these systems, although simpler to make than solid ones, are tremendously difficult to adjust and very difficult to operate. There is reason to believe, however, that the use of composite objectives is one of the trends which will be followed in the design of astronomical telescopes of the future, or at least that the idea will be used to solve problems which cannot be tackled in any other way.

How do matters stand with microscopy? On the one hand, conditions are more

favorable - the movement of the atmosphere does not effect miroscopes for obvious
reasons. Furthermore, (methods of illuminating the object are available and accordhg
(MANY written above "methods")
to circumstances we can chose the most suitable system. But here the root of the
trouble is the object itself. Structures vary; some objects may be gelatinous and
may refract rays passing through them (transparent gelatinous matter acts as a lens
and adds still more interference); some of the objective may have a fine, sharp
structure also causing diffraction. As a result there is so much distortion that
interpreting the pattern becomes an extremely difficult job. Nevertheless, the
problem can be fully solved. Much has been done along these lines by Rozhdestvenskiy.
We can hope that new, fresh scientists will take the trail already blazed, (In fact
(WHICH IS written above)
the only right one, despite all its difficulties.

Another obstacle which is often encountered in microscopy is poor contrast
of images. It is due to the fact that the object consists of sections only slightly
different in color, on account of which the boundaries between them are hardly
distinguishable.

There are many methods for improving contrast, among which we will point out
the system used by Academician V. P. Linnik (the rays pass through the object twice,
thus increasing the contrast) and "color" photography in the ultra-violet region of
the spectrum, developed by Ye. N. Brumberg at the laboratory run by Academician
S. I. Vavilov.

The second method is as follows. Three photographs are taken of the same
part of the preparation at different wavelengths using a microscope transparent to
ultra-violet rays and equipped with a monochromator for the ultra-violet region of
he spectrum (this makes it possible to change the wavelengths of the light illuminating
the preparation). The photographs, either in the form of negatives or printed
positives, are projected onto an area of a screen by three projectors, in front of
which there is a light filter transmitting only one of the three basic colors - red,
green or blue. The screen shows a color picture, the tinting of the elements in
hich is determined by the difference h the blackening of the corresponding points
on the photographs. The coloring, however, is conditional, though it is related to

the nature and chemical composition of different parts of the preparation. The colomed picture not only possesses greater contrast than the normal black and white but also provides useful indications of the distribution of different substances in the preparation and to some extent even acts as a substitute for chemical analysis.

Fig. 47 shows a "colored" ultra-violet photograph of a cross-section of a pinecone, taken on wavelengths of 400, 300 and 250 millimicrons. Green, red and blue filters were used.

Another method of improving contrast, based on the use of phase plates, will be described in Ch. V.

Great assistance can be rendered in recognizing the shape of viewed objects, particularly their depth, by the stereoscopic effect used in binocular microscopes; this effect enables us to see the image with both eyes. Another method of solving the problem was put forward by Linnik. It is based on interference of two beams. In the Linnik microinterferometer the observer sees the image of the object against a background of fine hairs - lines resembling a topographical map with horizontals.

At points where these lines intersect with the object, they form planes parallel to one another half a wave apart, i.e., 0.3 microns apart. They represent sort of horizontals indicating the relief of the object. Fig. 48 shows a photograph of an interferogram. It has been taken from the end measure (Joganson plate) at the stage of incomplete processing. It is easy to determine the depth of the lines from tortuosities. The distance between two black and white lines representing half the wavelength of the radiation is taken as the depth scale.

The photograph clearly shows numerous scratches, the depth of which lies within several tenths of a micron.

This method provides valuable information on the structure of diatom scales and has been successfully used to clear up a number of dubious facts regarding the structure of folds in the diatom Pleurosigma angulatum (W. Smith).

Another example of the use of the interference method is the treatment of the photograph of the diatom Staurineis.

On the left-hand sand of Fig. 49 we reproduce a photograph of the shell of

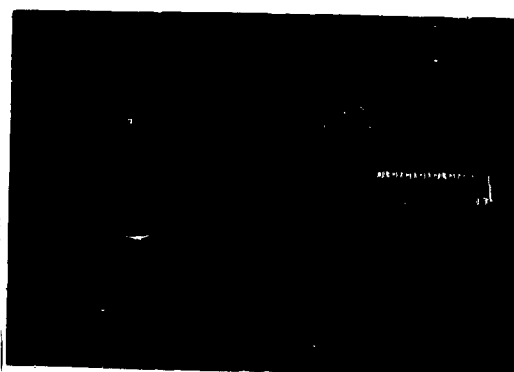Fig. 47. A "colored" ultra-violet photograph of a cross-section of
a pine cone.

Fig. 48. Interference pattern AN END measurement
AFTER LINNIK.

107

this diatom, taken through an ordinary microscope. The interference pattern cannot be reproduced in print on account of the conglomeration of detail, hence we only show a few interference lines, the most interesting ones, along side the photograph of the diatom.

The horizontal, 2, reveals the presence on a smooth field occupying the middle of the shell of several surface twists completely invisible during normal analysis. Furthermore, the concavity of the curve facing upwards indicates that the field, despite all obviousness, rises above the rest of the shell surface. The zig-zag nature of the longitudinal curves, 4, 5, and 6, indicates the presence of transverse ribs, also invisible in the left-hand drawing. Curves 1 and 3 give an idea of the cross-section of the shell; the black longitudinal line is not a cavity, but a protuberance. The horizontals provide a lot more, very valuable information on the structure of the diatom.

It should be pointed out that the interferogram method was first used for biology in our country and has already produced a number of interesting new results.

This does not exhaust the possibilities of delving into the region of the infinitely small. But among them there is no place for supermagnification, since this method carries us further away from correct images, AND (in effect DEPRIVES US OF ALL VISION since the pupil in this case is cut down to such a small orifice that our eyes are not in a position to distinguish anything on account of diffraction.
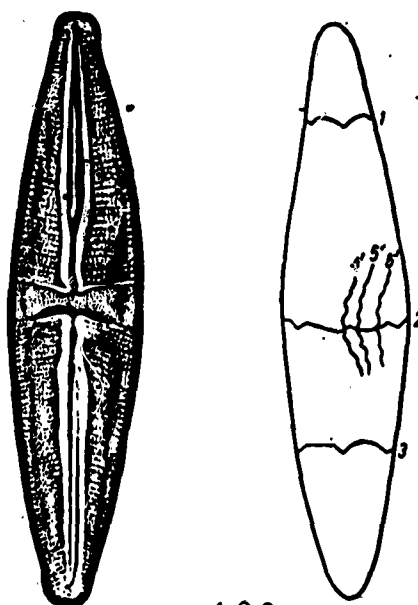

Fig. 49. Photograph of armor of diatom Staurineis

108

Sec. 9. Limitations imposed upon optical systems by aberration

In the foregoing chapters we have disregarded the effect of aberration on the quality of the image. The possibilities of optical instruments are limited either by the undular nature of light, which restricts the enlargement of optical systems, or by the general laws of optics such as the Lagrange-Helmholz law, proving the impossibility of arbitrary modification of the structure of beams, and, in particular, spelling failure for all attempts to set light to objects a long way away.

The aberrations inherent in optical systems, for their part, also impose limitations on the characteristics of optical systems, for example aperture ratio, angle of vision, and resolving power. Although aberrations, as distinct from diffraction effects in the IMAGE at a point, can be theoretically corrected by increasing the complexity of the design, in actual fact there are obstacles which in practice prove unsurmountable. Few people know about these. Even those who use optical instruments usually under-rate the difficulties involved in crossing the boundaries imposed by the contemporary state of the theory of optical systems, and insist that designers meet absolutely impossible specifications.

Let us deal only briefly with the limits of the principal characteristics of telescopic systems, photographic lenses and microscopes. In telescopic systems the angle of vision of the eyepiece is limited; for the normal wide-angle eyepieces it cannot exceed 60 - 65°, and in the case of the very complex designs developed at the State Optical Institute over the last few years it is 100 - 110°.

This conditions the objective field of vision. Indeed, if the enlargement of the telescopic system is γ, the angle of vision in the space of the objects is 2w, and in the space of the image 2w', then

$$\operatorname{tg} w = \frac{\operatorname{tg} w'}{\gamma}.$$

The relative aperture of astronomical refractors consisting of two lenses cannot exceed a certain point on account of the so-called secondary spectrum, i.e., the residual (irremovable) chromatic aberration which stems from the fact

that the frequent dispersions of the objective lenses are not strictly proportional to each other. To avoid excessive aberration rings due to this aberration, at focal distances of the order of 1 m, the objective can be made to provide relative apertures of up to 1:10 at focal distances of 2 - 4 mm up to 1:12 - 1:14, and, finally up to 1:15 in the largest refractors.

The use of certain materials such as lithium fluoride and fluorite, which are now made artificially, will enable an appreciable increase to be made in the relative aperture of these objectives. The same thing could be attained by using phase layers coated on the surface of one of the objective lenses.

It is curious that double or treble apochromatic lenses, which are made of a particular type of glass, cannot be made with the relative apertures mentioned above for ordinary double objectives on account of the great curvature of the surfaces which causes a considerable residual spherical aberration; at focal distances of 1 - 1.5 m, the relative aperture should not exceed 1:15 for treble apochromats, and 1:12 for double apochromates.

Eyepieces are not designed for relative apertures usually exceeding 1:4 - 1:3, mainly on account of the aberration of inclined beams (coma, astigmatism).

Microscopic objectives, particularly those providing high magnification (from 60 to 100) should have a large numerical aperture A, providing good resolving power. Hence when calculating them, particular attention should be given to correcting the aberration of the center of the image, i.e., to spherical and chromatic aberration. The quality of the image very rapidly deteriorates as the point of the object moves further from the microscope axis. It can be considered that the linear field of vision of strong microscope objectives is not more than a twientieth of the focal distance (f)

$$2l < \frac{f}{20}.$$

If β is the linear enlargement of the objective, its focal distance is determined from the equation

$$f = \frac{160}{\beta},$$

where 160 is the optical length of the tubus in millimeters. Taking this expression

1 1 0

into account, we get for the linear field of the microscope

$$2l < \frac{s}{\beta}.$$

For low-magnification objectives this equation is no longer valid since objectives of the photographic type can be designed; these give greater fields.

At the beginning of the century there began to appear microscope objectives with a corrected surface curvature, which contained extra, thick meniscus-type lenses, as distinct from the conventional type. The linear field of vision of these objectives was greater than in ordinary ones, and they were particularly suitable for photographing greatly elongated objects.

Mirror-lens objectives have begun to appear recently, also corrective for curvature of the image and for chromatic aberration. Unfortunately, the inestimable presence of central vignetting somewhat spoils the quality of the image produced by these objectives in that it strengthens the brightness of the diffraction rings and thereby reduces the contrast. The use of these objectives may be extended to the ultra-violet region.

Photographic objectives are used by an enormous number of amateur photographers and quite a few specialists in a variety of professions. It is no wonder that strict and often contradictory demands are made upon these objectives, for example, large aperture ratio, considerable field of vision and a high resolving power, to boot. The design, moreover, has to be simple, light-weight and there should be no loss of light. Naturally, all these conditions are incompatible and it is only the specialized objectives which are good.

An ultra-rapid objective with a relative aperture 1:1.5 intended for bad lighting conditions, with a focal distance of the order of 100 mm, cannot have an angular vision greater than 35 - 40°. When the focal distance is 50 mm, it may provide fairly good images on a normal, double frame (24 x 36 mm). A reduction in the relative aperture K, keeping approximately the same degree of complexity in design, increases the angle of vision 2w, as shown in Table 2 calculated for a focal distance of 100 mm.

It can be seen from Table 2 that since the first edition of this book (1944)

the angle of vision 2w for K equals 1:1, 1:2, 1:3 has become 5 - 10° larger. This increase for the same relative aperture shows the advancement in theoretical optics over the last decade. Apart from improvement of the quality of the objectives (better corrected than the earlier ones) we should point out the production of a number of new objectives, for example, an increased rear segment, intended for mirror cameras, or objectives for color photography, and so on.

Table 2 gives the actual limits of present-day photographic lenses of medium complexity. Specifications which do not lie within these limits either result in a poor image or else excessively complicated designs.

The phenomenal development of mirror-lens systems has made it possible to reduce the size and number of optical instruments, which is particularly important in astronomy and "long-range" photography, which require large diameters and only slight weight (a considerable advantage in transporting the objectives). The most difficult thing to achieve here is a small ratio between the length of the tube and its diameter (rather than the focal distance which might have seemed natural). At the present time the Soviet optical industry is producing mirror-lens objectives, the length of which is less than the diameter.

The use of aspheric surfaces has enabled us in certain instances to simplify the design of the optical system by reducing the number of lenses but without any deterioration of the quality of the image. The manufacture of these surfaces has not yet been assimilated with a high degree of accuracy.

Calculations show that by using combinations of parabolic and hyperbolic mirrors and small correction lenses the length of the objective can be made 3 or 4 times shorter than its diameter.

Certain progress has been made in constructing extremely rapid photographic lenses by using "super-heavy" crown glass, the composition of which includes lanthanum and other allied materials. Back at the beginning of the century methods of manufacturing artificial crystals were poorly developed and the optical industry hardly produced anything but lithium fluoride, whereas now dozens of crystals are being produced, including fluorite and a number of other types transparent in the

Table 2

K	2w, degree	
	1944	1957
1:1	15	20
1:2	40	50
1:3	60	60—65
1:4,5	75	75
1:6	90	90
1:8	120	120
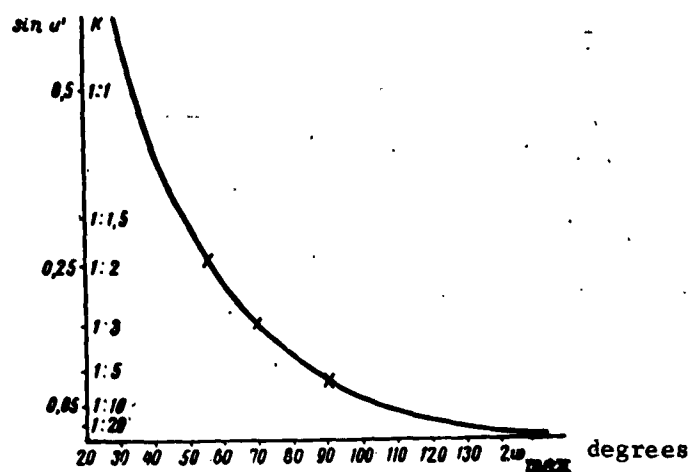
Fig. 50. Limit of possibilities of modern camera lenses (f = 50 mm);

113

infra-red region and opaque in the visible region of the spectrum. All this has greatly stepped up the chances of making spectral devices, in which these devices can be used.

In every group of optical instruments there is a certain ceiling in their basic characteristics: the relative aperture, angle of vision and size. The dependence between the aperture (relative aperture) and corresponding maximum angle of vision stands out particularly.

The relationship between the relative aperture K and the maximum angle of vision for high-grade objectives used for different purposes but possessing the same total distance is shown by the smooth curve in Fig. 50. The curve represents the limit in the possibilities of present-day photographic lenses. The limit shifts slightly with time and rises as the design of optical systems becomes more complex, new materials are produced, and so forth; but, the shift is a slow one and is evidently tending to a certain limit. What is the reason for this phenomenon, which has been known for a long time but not been given any explanation? The answer, of course, lies in the theory of aberration, in the section dealing with residual aberration of a high order, for the moment inaccessible to us.

It is curious that the product l' sin u', in which l' is the distance between the image point and the axis of the system, and u' is the aperture angle in the image space is constant along the boundary curve. The latter is a curve close to a hyperbola. It is strange that this product is exactly the expression contained in the Lagrange-Helmholtz law. But it is completely incomprehensible why it should be just this invariant, which has no bearing on the aberration of the optical system, but merely relates to the luminous energy flux incident on the film, that is the same for all medium-complexity objectives (or highly complex ones).

Sec. 10. The X-ray microscope

The wavelengths of electromagnetic vibrations in the visible region of the spectrum range from 0.3 to 0.8 microns, hence the least resoluble distance shown by normal optical systems is of the same nature.

Much greater possibilities, at least as far as theory is concerned, can be

expected from systems operating with x-rays, the wavelengths of which range from 1 to 100 Å (1 micron equals 10 Å), i.e., three orders smaller. Hence after Laue's experiments making it possible to determine the wavelength of these rays (1912), creative thought began to seek ways of constructing an x-ray microscope. The job proved a tough one. Roentgen's rays are hardly refracted at all and only are reflected when they graze a surface, hence conventional optical systems are here unsuitable.

A simpler system suitable for studying very thin layers of matter is called the contact system. Its principle is as follows: a thin layer of the substance being tested (0.1 - 0.2 mm) is brought into contact with a fine grain (high-resolving) photographic plate, after which the layer is irradiated with a beam of x-rays, as far as possible parallel. Since the plate is sensitive to these rays, when it is processed, it shows a full-size image of the layer. The image is then enlarged optically.

A disadvantage of this method is its unsuitability for low-absorbing layers. The use of thicker layers leads to blurring of the image and a loss of resolving power on account of the non-parallel incident beam. It should be pointed out that the resolving power is not so much restricted, by the wavelength of the x-rays as the grain size of the emulsion on the plate.

Another version of this "shadow" method is the following. A point source of x-rays emits a beam. The object to be observed is placed a short distance away from the beam in the form of a very thin section, magnified shadow of which is projected onto a screen some way away from the source. To attain a point source there are various methods, but they all boil down in the long run to producing a sharply focused electron beam at the anode by means of an electron microscope (operating in reverse). The path of the anode bombarded by electrons forms a virtual point source of x-rays.

Fig. 51 shows a diagram of a Soviet device in which the point source is formed by an EM4 electron diffraction camera*.

----
* For more details see article by G.O. Bagdyk'yanets. X-ray shadow microscope. Biophysics, 1957, 1, Vol. 4, p. 341.
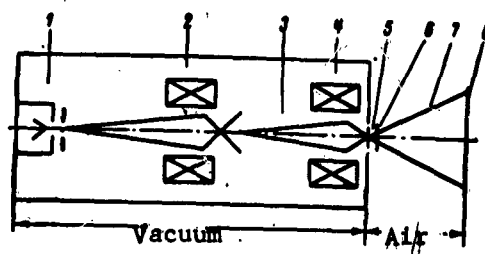
Fig. 51. Diagram of x-ray shadow microscope in which point source is formed
by a diffraction camera. 1) electron source; 2,4) reducing lenses;
3) electron beam; 5) anode; 6) specimen; 7) x-ray beam; 8) film.

It should be pointed out that on account of a number of technical diffi-
culties, the least resoluble distance of x-ray/microscopes is not less than 1 - 0.5
SHADOW
microns for the moment. No great reduction in this figure is expected in the near
future.

Thus, x-ray shadow methods are inferior to optical ones from the point of
view of resolving power, and were it not for a number of advantages compared with
conventional optics (for instance, the possibility of penetrating dense, opaque
bodies), they could never compete.

The microscopes closest in nature to conventional "visual" ones are mirror
microscopes consisting of reflecting surfaces. As shown by experience and borne
out by theory, the refractive index of metal media through which x-rays pass is
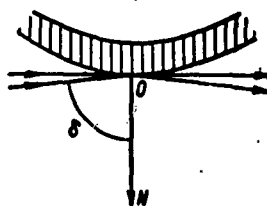very close to unity (slightly less than unity).

Fig. 52. Total "external" reflection of x-rays.

Hence, instead of the normal total internal reflection during refraction
through media transparent to x-rays, we find "total external reflection". What
happens here is that the rays which form an angle greater than the limiting

116

with the normal to the surface at which reflection occurs are totally reflected (Fig. 52), the angle being determined from the equation

$$\sin \delta = 1.64 \cdot 10^5 \, \lambda \sqrt{\rho}.$$

where ρ is the density of the medium in g/cm^2 and λ is the wavelength in cm.

Since the difference between the refractive index and unity is only a few ten thousandths of unity, the angle δ, as shown by calculation, does not differ from a right angle by more than a few degrees. Simple refraction in this case is so weak that it requires tens and hundreds of lenses of the most refracting metals to produce the magnifications attained in conventional visual light microscopes.

Hence x-ray microscopes only make use of total external reflection, and the mirrors are placed at extreme angles to the incident beams. A special geometric optical system has to be built, in which the angles of incidence and reflection are close to 90° and a method of removing aberration has to be found, the main defect in which is the absence of a paraxial region (a region in which the angles of incidence and reflection are close to zero).

Under these circumstances there is no image in the conventional sense, and we have to use special measures by selecting the shape and position of individual surfaces so that a point is represented by a point, even though by means of infinitely fine beams. Use is therefore made of separate elements (Fig. 53) consisting of two cylindrical mirrors placed perpendicular to each other. It is quite difficult to see to it that the plane of the image is perpendicular to the principal (central) ray in the beam. Finally, it can be shown that correction of the aberration "coma" is impossible. This means that the aperture of the rays emitted from the objects does not exceed 0.05 - 0.1.
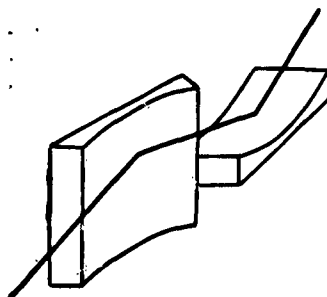
Fig. 53. Reflecting elements in an x-ray microscope.

117

Furthermore, the accuracy required in manufacturing surfaces of cylindrical shape on account of the low wavelength of x-rays has to be dozens of times greater than the accuracy required for the most highly precision optical mirrors. Even in ideal manufacture, because of the above difficulties (low aperture, coma), it is impossible to bring about a great increase in the resolving power, compared with the power provided by ordinary "visual" microscopes - instead of the theoretically possible increase of the order 10^3 - 10^4, is hardly possible to count on an increase of 10 or 20, and even then tremendous difficulties are involved. Nevertheless, a few models of microscopes of this type have been built.

X-ray spectrographs, which make it possible to carry out spectral analysis over the range 1 - 100 Å, are of still greater importance in engineering than x-ray microscopes.

In this case prisms cannot be used as the dispering systems, since the refraction is negligible, even for the densest metals. Ordinary refraction gratings are no use either (except for the difficult and useless case of rays striking the lattice plane at extremely small angles of the order of 10 - 20') since the distance between the lines has to be of the same order as the wavelength of the rays used, and the manufacture of gratings with millions of lines per millimeter is inconceivable, or at least at the moment.

But nature provides us with natural gratings: crystals constitute an aggregation of atoms in the form of a regular spatial lattice, the distance between the atoms amounting to 1 - 5 Å. Each atom in the crystal as a scattering center for x-rays which are coherent among themselves. These rays interfere with each other and exhibit maxima in different directions. If the crystal (mica or plaster of paris) can be made cylindrical by bending, it acts as a concave grating and an x-ray spectrum of the material being studied can be produced on light sensitive plates (Fig. 54).

X-ray microscopes have been of tremendous practical importance, since they have enabled us to delve deep into matter and see things which cannot be observed in any other way. But they have not solved the principal optical problem, namely

increasing the resolving power in comparison to ordinary "visual" microscopes. As is well known, the least resoluble distance in the matter is 0.2 - 0.3 microns.

X-ray microscopes have not provided this degree of resolution for the moment. Each of the above-described types of x-ray microscope has a particular reason for this. In contact microscopes the reason is the insufficiently small size of the source; in microscopes consisting of mirrors the aperture is too small and it is difficult to make the surface with the required precision.

In their search for new ways of increasing resolving power, inventors have turned to beams of electrons, the wavelength of which is still several orders smaller than those of x-rays.
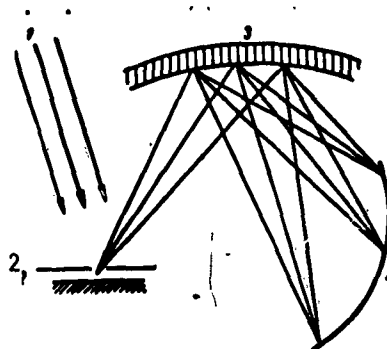

Fig. 54. Formation of x-ray spectra: 1) electron beam; 2) x-ray source; 3) crystal; 4) light sensitive layer.

Sec. 11. Electron microscope

Over the last two decades it has become possible to use non-luminous radiation for purposes of microscopy, that is to say radiation the wavelength of which may be thousands or millions of times smaller than for light.

A beam of electrons possesses these properties. It cannot, obviously, be observed directly by the eye, but when they impinge upon a fluorescent screen, electrons cause fluorescence which can indeed be observed; the beam of electrons also leaves a trace on a photographic plate. There is a far-reaching similarity between light radiation and an electron beam.

In both cases a certain factor (rise in temperature, illumination and so

forth) cause the object to radiate particles - light quanta in the first case and electrons in the second. In their motion they are both equivalent to a train of waves, the length of which is a function in the first case of the properties of the radiating object itself, in the second of the flight velocity of the electrons. The flight velocity may be varied over a wide range by superimposing an accelerating potential difference (for example, in electron/radio tubes). The relationship between the wavelength(in Angstroms) and the applied potential difference V in volts is shown by the equation

$$\lambda = \sqrt{\frac{150}{V}} \, .$$

For instance, if the potential difference is 150 v, the electron wavelength is 1Å, i.e., 5000 times shorter than the wavelength of light. At values of V of the order of 10^5 v, commonly used at the present time, the wavelength $\lambda = 0.04$ Å, i.e., 10^5 times smaller than in an optical microscope.

In the same way as light waves which are refracted when passing through media with different refractive indices, the electron changes its direction under the influence of an electric or magnetic field. The fields act as lenses refracting the path of the light rays. The laws of the refraction of electrons follow from the Fermat principle exactly in the same way as laws of the refraction of light, and for this reason the general formation of images in optical systems can be applied without modification to electronic-optical systems. Not only do the laws of paraxial optics coincide, according to which the image of a point is a point, rather than a straight line, and so forth, but electronic lenses also produce similar aberrations, and the aberrations (to a much greater degree than in optical systems) restrict the resolving power of the systems.

We cannot deal here with the technical aspect of electronic-optical instruments, especially the calculation and design of electronic lenses. For what follows we need only know that by selecting the right shape, number and arrangement of the terms and metal parts creating the magnetic field, or condenser casing creating an electrostatic field, we can satisfy the following conditions: with a certain degree of accuracy the refraction of the electron path is proportional to

the distance between the point of the trajectory and the axis of the system - a property possessed by optical lenses.

The magnetic field causes the electron to deviate both in its direction of motion as well as the direction perpendicular to it, so that when passing through a magnetic lens field it describes a corkscrew with a variable radius (Fig. 55). But this fact does not effect the power of the electronic-optical system to produce images, providing this condition is satisfied.
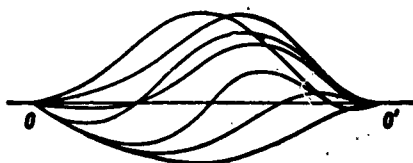

Fig. 55. Paths of electrons in magnetic field.

Thus, the behavior of an electron beam (at least schematically) can be described in exactly the same way as an optical system, and a simplified diagram of an electron microscope is an exact reproduction of an optical one.

In the same way that in an optical microscope the object viewed may be self-luminous or lit by an outside light source, in the electron microscope we can use the electronic emission of the object itself produced by either heating it or by the action of light rays (visible ones, ultra-violet or x-rays), or else by the action of ions or electrons, or alternatively by passing an electron stream through the object, which absorbs the latter to a lesser or greater degree and produces an image.

To boost the speed of electrons radiated by the object or by an outside source, a potential is created at the anode differing from the potential at the cathode (electron source) by tens or hundreds of kilovolts. The beams of electrons emitted by the object are gathered by the first objective lens in the plane of the primary image and by the second lens of the objective on a fluorescent screen or photographic plate.

121

Fig. 56 is a diagram of the electron microscope with magnetic field. Electrons leave the cathode, 1, of the discharge tube or an incandescent filament; the high potential (with respect to the cathode) applied to the anode, 2, boosts the electrons, and some of them pass through the small opening drilled in the anode. The electrons are concentrated on the object by the magnetic lens, 3, which plays the part of a condenser. When they reach the object, 4, the electrons pass through it and are ABSORBED to a lesser or greater degree or else scattered according to its structure, and after being twice refracted by the objectives, 5 and 6, form a greatly magnified image of the object, 8. The image is either viewed directly through the observation window, 7, on a fluorescent screen, or else photographed.

A vacuum of the order of 10^{-4} mmHg is maintained inside the electron microscope tube; the presence of air, even though greatly rarefied, causes considerable absorption and scatter of the electrons. The same effect was produced in the ordinary microscope by matte glass, or more accurately, a number of matte lenses placed in the path of the rays.

An interesting feature of electron microscopes is the possibility of varying the magnification over a wide range with hardly any change in the position of the lenses, by varying the current passing through the lens windings. The useful enlargement of the electron microscope is now as much as 100,000.

Let us now turn to a problem of the greatest interest to us, i.e., what we can see by means of this type of microscope.

The matter is by no means as simple as in the case of a microscope using visual light, the theory of which has been fundamentally worked out during the 300 years of its existence. The electron microscope has been developed in very recent times and for the moment there is not much data available on it. To forecast its possibilities is almost as difficult as it was during Levenguk's time to judge what microscopes would be like at present. Nevertheless, the experience which we have gained in the use of optical microscopes and the information which is
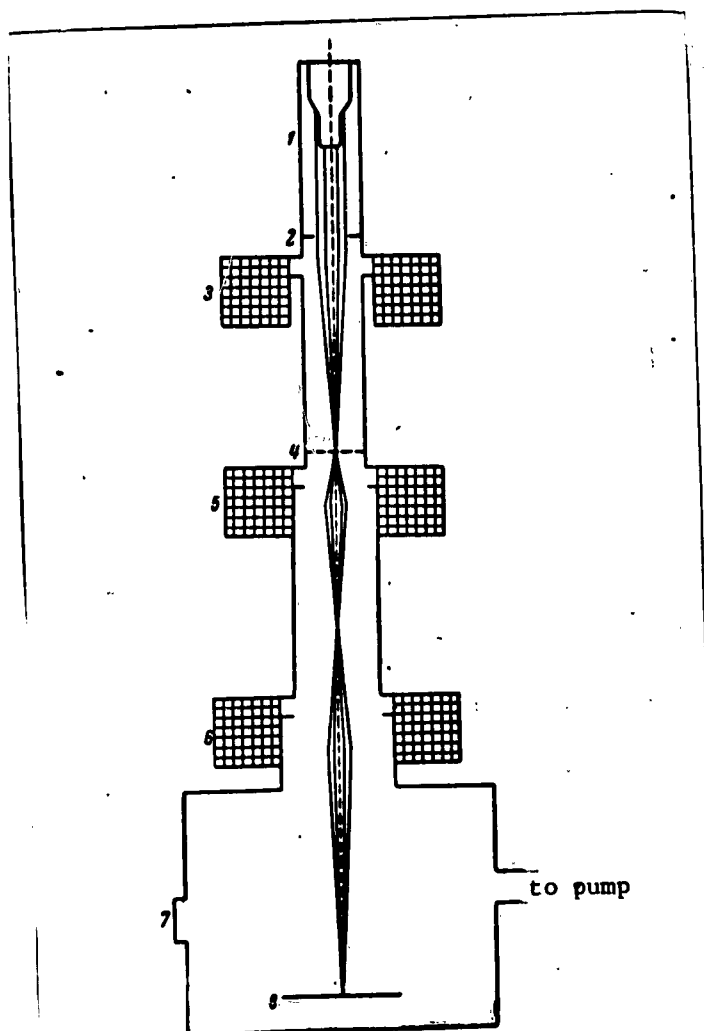
Fig. 56. Diagram of electron microscope with magnetic field.

provided by the present-day electronic theory enable us to make one or two pre-dictions.

The possibilities of any instrument are limited, on the one hand, by the perfection of design and, on the other, by the principles on which it works.

The design of a present-day optical microscope is so far advanced that further improvement cannot bring about any increase in resolving power, the limit of which is determined solely by the wavelength of light.

The electron microscope is still in its infancy. Its lenses are probably even inferior in properties to those used by the first microscopists in the XVII century. The elimination of aberration in the electronic lenses is by no means an easy task and the methods of correction are unclear. Furthermore, certain aberrations, for example, chromatic (due to varying electron velocities) are evidently irremovable in principal, since when passing through the structure of the object, the electrons lose velocity to some extent at different points.

The only way of reducing aberration is to decrease the aperture of the microscope, since aberration is inversely proportional to the degree of the aperture angle. At the present time the numerical aperture of electron microscopes is not more than 0.02, whereas for optical microscopes it can be as much as 1.5, which is 100 times greater.

We need only recall, however, the development of the optical microscope, in particular Newton's statement of the fundamental impossibility of correcting chromatic aberration, to realize the need for the greatest caution in making comments of this kind. It can be assumed that in time means will be found of dealing with the aberrating and increasing the aperture.

A defect of the electron microscope difficult to remove in principal is the mutual repulsion of the electrons, as a result of which they are unable to gather in the plane of the object at one mathematical point. But this disadvantage can be reduced to a minimum by electron streams of very low density not apparently limiting the resolving power of the microscope.

The wavelength of vibrations caused by the electrons may be reduced to

hundredths of an angstrom, as shown above. Hence on an analogy with the optical microscope, we should expect the least resoluble distance to be of the order of hundredths or tenths of an angstrom, i.e., less than the size of a molecule or atom. (The diameter of a molecule ranges from several angstroms to hundredths of an angstrom for molecules of certain proteins).

In actual fact, the limit of the resolving power is set by the nature of the electron and its action on the object which it is supposed to detect. An electron can only detect an object, for example, a molecule, if it collides with the latter. But the mass of a molecule is not really so large, compared with that of the electron, and when the latter is moving at a tremendous rate, the molecule either loses its individuality and is destroyed, or, at best, is discarded from its original position at a rate of about 1 km/sec and immediately disappears from sight. It is only the giant molecules mentioned above which can be observed on account of their great inertia.

According to the Heisenberg UNCERTAINTY relationship, the less the mass of the element of matter, the more UNCERTAINT its state of motion, i.e., its coordinate and velocity. This UNCERTAINTY is already so large for molecules that to see them in motion (in the same sense that we see the motion of an object with the naked eye) is quite impossible. There is, however, a connection between the Heisenberg relationship and the views expressed above on the collision between the electron and molecules.

Thus, the limit for viewing through an electron microscope can be considered large-size molecules; this limit is hundreds of times greater than the possibility of an optical microscope. It should be pointed out that over the few years that the electron microscope has been in existence, its resolving power has already become 30 - 50 times greater than in the optical microscope.

Unfortunately, the possibilities of an electron microscope are limited for another reason, namely, the electron beam impinging upon the object has an effect on it and in certain cases destroys it instantaneously. It is especially dangerous to bombard living cells with electrons since they are hardly in a position to with-

stand them with impunity. But, here, too, there are methods enabling this
difficulty to be overcome.

CHAPTER V

PHASE, AMPLITUDE AND IMAGE

SEC. 1. The role of phase in the wave theory of image formation

The action of an optical system may be fully explained by the undular nature of light. But none of the writers of textbooks on the theory of optical instruments has dared to do so since THEIR *REASONING* would abound in mathematical obstacles of too serious a nature. They usually limit themselves to proof of the fact that the wave theory of light expressed by means of Maxwell equations implies in the limiting case of infinitely small light wavelengths the well-known fundamentals of geometric optics, determining the path of a ray through a medium with a variable refractive index, in particular, the Snell-Debartes laws.

The further development of the theory of reflection by optical systems is based solely on these laws, and the function of the optical instruments, in its turn, is based on the general properties of systems derived from the laws of geometric optics. It is only at a later stage, when the student reaches the problem of the resolving power of an optical system, that he has to recall the finite wavelength of luminous oscillations and determine the true distribution of intensity in the image of a point produced by an ideal (non-aberrating) optical system. Here geometric optics is not left out, since it is still required for determining the wave surfaces - a concept belonging entirely to geometric optics, but essential as well for calculating defraction phenomena.

This method of description, which is probably the only possible one for didactic reasons, is apt to lead to the view widely held by THE MORE *INEXPERIENCED* specialists in optics that in general outline an image is produced in accordance with the laws of geometric optics, while the wave nature of light, which in the given case shows up in diffraction phenomena, only changes the picture of the image to a slight degree. For practical purposes the change is only appreciable when the object in question possesses an extremely fine structure (a point, line, grid and so on).

The image of a point produced by a non-aberrating optical system takes the

form of a point in accordance with geometric optics, and as a result of diffraction phenomena, takes the form of a tiny spot surrounded by feeble rings.

In the more general case of a conventional optical system with aberrations, the effect of the wave structure of light is only that the aberration pattern, which hardly changes at all in dimension or general appearance, assumes a more intricate nature, accompanied by very thin light or dark lines that sometimes create complex attractive patterns. In short, direct observation of images of objects with a fine structure produces the impression that diffraction only redistributes the luminous energy to some extent within the image of the object, little affecting the image pattern as a whole. It can further be said that the wave structure of light introduces only slight corrections to the image governed by the laws of geometric optics: the corrections can be disregarded in all cases in which the image is not studied at an advanced level.

When we are concerned with the image of uniform areas of more appreciable dimensions (depending on the resolving power of the optical system), rather than a point or an object of fine structure, we no longer find any fine structure in the image and the pattern of the latter assumes a form in accordance with the laws of geometric optics. In this case the diffraction almost totally disappears.

These facts explain the conventional arrangement of information in textbooks on geometric optics and the theory of optical instruments, which the beginning of the course deals with the theory of images based on the simple laws of geometric optics, and it is not until much later that we find the chapter on the diffraction image of a point, and even then it is sometimes in smaller print or contained in a supplement which need not necessarily be read.

It can be easily shown that the view of the part played by the wave structure of light IMPLICIT IN this arrangement of the text is radically wrong. Readers gain the impression that when there is a slight variation in the optical system completely unaffecting the geometrical pattern of the image (for example, when THE thickness of part of the surface is enlarged by a magnitude undetectable by the eye), there need not be any appreciable change in the actual image. In actual

128

fact is may happen that the image becomes unrecognizable, and in particular, that it may vanish completely.

Sec. 2. Huygens -Fresnel principle

From now on we shall make frequent reference to the Huygens -Fresnel principle which we encountered en passant further back (p. 88). Let us now describe this principle in greater detail since it enables us to calculate the distribution of energy at different points in the beam of light with a fair degree of accuracy.

Let the point source S (Fig. 57) radiate rays normal to the spherical wave surfaces, one of which is AA_1. After refraction in the optical system OO', the rays remain orthogonal with respect to the new family of wave surfaces, among which we should note the arbitrary surface A'A, in accordance with the Malus theorem. Generally speaking, the surface is no longer spherical on account of aberration, poor quality of manufacture, or other factors. The surface is determined from the following condition: the optical paths Σ nl, i.e., the sum of the products of the segments l counted off along the rays and the corresponding refractive indices n are the same for all rays. This means that the light travels from point S (or the surface of the wave AA_1) to the wave surface $A'A'_1$ in the same amount of time for all rays (a corollary the Fermat principle).
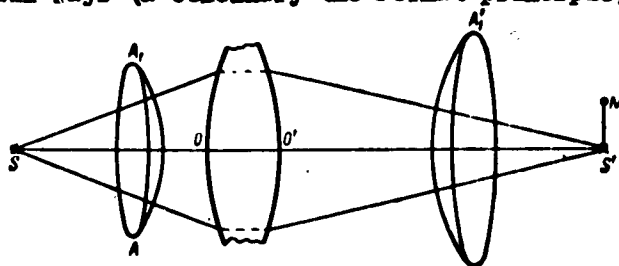

Fig. 57. Wave surfaces of rays incident on an optical system and refracted by it.

Knowing the position and shape of one of the wave surfaces in the space between the images, we can use the Huygens-Fresnel principle to determine the illumination of any point M in the space of the images. The illumination is equal to the square of the amplitude of the resulting oscillatory motion at point M, which, according to this principle, is defined as the sum|total of all oscillations originating from all elements of the wave surface AA' and reaching point M. The

effect of each element ... e ... ce A'b' is determined by the equation

$$\text{dscos}\frac{2\pi r}{\lambda},$$

where a is the amplitude on the ... rface A'b', ds ... area of the element of the surface ... the distance ... the ... ment ds and point M and λ is the wavelength.

The point of this ... the elementary pulse received at point M is proportional to the ... and inversely proportional to the distance r, as indeed follows from the ger ... law of the propagation of energy. Fresnel's idea is represented by the multi ... $\frac{2\pi}{\lambda}r$, which expresses the time wave distribution of energy along the ray ... ing the ... element ds to point M at a moment of time t. We ... not include the ... me t ... ... vanishes during the calculations. The coefficient ... is re ... to the brightness of the beam, which in turn is governed by the brig ... tness of the source itself.

Thus, the effect of each element ... point M can be represented as the product of two quantities. The first multipl ... ds varies very slowly as the position of the element ds on the wave surface ... hanges; this is the amplitude at point M. The second multiplier $\cos\frac{2\pi r}{\lambda}$, which ... resents the interference of the elementary beams in mathematical form, ... ly with the change in position of the element ds; the angle $\varphi = \frac{2\pi r}{\lambda}$ ... ... the ... in phase as point M.
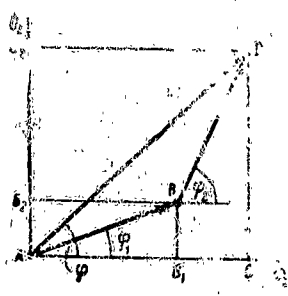

Fig. 58. Vector addition of light oscillation amplitudes.

Every impulse emitted by the area element ... ds, be represented ... by the vector AB (Fig. 58), the amplit ... of which is ... ... which forms ... an angle $\varphi = \frac{2\pi r}{\lambda}$ with the axis

The amplitude resulting ... ... the addition ... ... is

obtained by vector addition of the component impulses. In other words, the total amplitude is represented by the sum of all vectors similar to AB.

Let us recall that the vector representing the sum of the two vectors AB and BC is the vector AC joining the origin of vector AB TO the end of vector BC, the beginning of which adjoins the end of AB. We should point out that the projection AC_1 of vector AC onto the axis AO_1 is

$$AC_1 = AB \cos \varphi_1 + BC \cos \varphi_2.$$

where φ_1 and φ_2 are the respective phases of the vectors AB and BC. The projection of AC_2 of this vector onto the axis AO_2 is

$$AC_2 = AB \sin \varphi_1 + BC \sin \varphi_2.$$

The square of the amplitude of the resultant vector AC, i.e., the illumination produced at point M by the two elements ds_1 and ds_2 is equal to the sum of the squares of the projections, i.e.,

$$E = (AB \cos \varphi_1 + BC \cos \varphi_2)^2 + (AB \sin \varphi_1 + BC \sin \varphi_2)^2.$$

This result may be extended to cover any number of elementary vectors, representing impulses from different elements of the wave surface. As an example, let us consider the following case: let it be assumed that part of the optical surface bounding an optical medium protrudes above the rest of it by the quantity Δ , determinable from the condition that the difference between the path of rays passing through the protruding part and rays passing through the rest of it is HALF THE WAVELENGTH. Furthermore, let is be assumed that this part occupies half the area of the surface. The difference

$$\Delta = \frac{\lambda}{2(n-1)},$$

where n is the refractive index of the medium which this surface separates from the air. If λ = 0.0005 mm, and n = 1.5, then Δ 0.0005 mm, i.e., the difference between the thicknesses is so small that it cannot be detected either by the eye or by touch. But the effect of this projection is so great that in accordance with the Huygens principle mentioned above, it totally eliminates the image of the point, or at least for the wavelength in question. Indeed, for every element of the area

131

of one part of the surface there is an element of the same area on the other part, and for every vector-impulse of one element of the projecting part there is a vector of the same length on the other part, but directed in the opposite direction, hence the resultant amplitude of the vector in the plane of the image (corresponding to a properly made surface) is equal to zero.

Here, however, the total disappearance of the image does not occur, since unless the light is monochromatic, which usually happens, rays with wavelengths differing from the basic one still take some part in the formation of the image.

This example illustrates the tremendous effect of variation in phase (or optical path) in a system on the type of image, even when the variation is due to factors so insignificant that they cannot be detected either by the eye or by touch. The fact has been known for a long time (it follows directly from the Huygens-Fresnel principle established at the beginning of the XIX Century), but it is only over the last few decades that it has come to be used in dealing with a number of problems involving the quality of an image produced by a system. Such problems include increasing THE resolving power, the removal of harmful FRINGS surrounding a diffraction image of a point, increasing contrast in the images of low-contrast objects, reducing chromatic aberration in optical systems and so on.

* *
*

Before proceeding to make a more detailed analysis of some of the applications of spherical properties resulting from the wave structure of light, we should point out once again how unfounded we consider the widely held view that optical phenomena involving the formation of an image may be explained in general outline by the laws of geometric optics, and that the wave structure of light only has a slight effect on the pattern of the image.

This conviction has mislead experienced specialists who are accoustomed to thinking in terms of geometric optics, and sometimes leads to serious errors, particularly in cases in which diffraction and interference are intensified, compared with what occurs in conventional optical systems.

No-one would obviously ever think of formulating the theory of the diffraction *grating* for the Fabry-Perot etalon on the basis of the law of geometric optics. But there are borderline cases in which it is not clear whether diffraction or interference is predominant. In such cases all calculations to determine the pattern of the image can only be made on the basis of the Huygens-Fresnel principle, and in those cases in which we deal with quantities much smaller than wavelengths (which admittedly is rare), not even the Huygens-Fresnel principle gives the right results.

Sec. 3. Amplitude filters

Apart from phase, the image of a point is Affected to a considerable extent by amplitude. We will leave aside the trivial case of an absorbant neutral light filter which weakens the whole image pattern a certain number of times, though it does not Affect the *relative* distribution of illumination and does not therefore Affect/contrast, either.

Let us consider a more interesting case in which the ENTRANCE PUPIL is fitted with a filter with an absorption factor differing according to area. It is not difficult to image how this filter Affects the image of a point in simple cases. For example, let us suppose the filter is a diaphragm which only lets light through a small central aperture. The image of the point (aperture) is BLURRED. The middle spot widens considerably and the diffraction rings spread out from the center. The resolving power of the system suffers greatly.

Let us suppose, conversely, that we cover the entire aperture with a diaphragm except for a narrow ring at the very edge. Both theory and practice show that in this case the central spot is reduced in diameter while the intensity of the diffraction rings is appreciably increased and approaches the intensity of the central spot. As a result the resolving power, classically defined, is increased, but on account of the brightness of the rings, the diameter of the image of the point is actually increased and the contrast decreased. A very odd effect is produced by the amplitude filter (Fig. 59) proposed by Soret (and also Rayleigh), which consists of alternating transparent and opaque rings of the same area, the outside radii of

133

which are proportional to the square roots of integers, exactly in the same way as for Fresnel zones by means of which Fresnel proved that the wave structure of light can be completely explained by rectilinear propagation. This filter, known as a zone plate, only lets through light during those phases which for a given position of the point object and its image results in intensification of the total amplitude, and do not let through light which has a weakening effect. Hence then it leaves an optical system, the ENTRANCE PUPIL of which contains a Soret plate, apart from the NORMAL image reduced to half its illumination, we can observe another series of images, which are strongly colored, though fairly weak.

But if the dark colored rings are made transparent and coated with a layer of substance creating a phase difference of half a wavelength compared with the remaining transparent spaces, as suggested by Wood, the basic image disappears since all the elementary impulses cancel each other out; we are left with "secondary" images which could be more correctly called the longitudinal spectra, on account of their extremely HIGH/chromatic aberration.

As we see, by MODIFYING the amplitude and phase of the oscillations passing through different points on the aperture, we can vary the image of the point over a wide range. The question naturally arises how wide are the limits? And in which spheres of optical engineering can we utilize these possibilities?
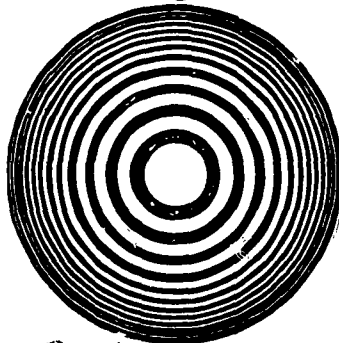

Fig. 59. Soret zone plate.

Sec. 4. Phase contrast in microscopy

When studying biological preparations under a microscope, observers often encounter the following impediment: transparent preparations (cells, bacteria, and so on) contained in a colorless and transparent medium are quite invisible through

134

the eyepiece, even when their dimensions considerably exceed the least resoluble [16] distance. In order to improve the visibility of the preparations, they are colored with special dyes, but the latter always disturb their functions to some extent and affect their behavior. The dyes very often result in destruction of the preparations, hence in many cases the coloring method is unsuitable.

A long time ago it was noticed that the microorganisms mentioned differ from their surrounding medium in refractive index (though, admittedly not to a very large degree). Furthermore, they often possess a "relief", that is to say they differ in thickness or depth   from the medium in which they are immersed. These organisms are sometimes known as "dephasing" since the light rays which pass through them undergo a phase jump discontinuity, compared with neighboring rays.

Beams of rays propagated through the surrounding medium and objects being studied are virtually unrefracted and move in exactly the same way as though the medium were homogeneous. Nevertheless, on the boundary of an object there is diffraction due to a small jump in the refractive index. The luminous energy is scattered to some extent or other, the scattering increasing as the particle decreases in size, and the discontinuity in the refractive index increases. Let us determine the magnitude of the phase jump. Let us assume that in a medium with refractive index n there is a transparent impurity (a particle M of thickness d - Fig. 60), the jump in whose refractive index is equal to Δ n.

Let us calculate the difference in the behavior of the two rays AA' and BB', one of which, the former, passes through the impurity, while the other (BB') only goes through the surrounding medium. We will take the latter to be bounded by two planes P and P', the distance between which is d_1. The optical path between planes P and P' for the ray BB' is

$$(BB') = d_1 n,$$

and for the ray AA'

$$(AA') = (d_1 - d) n + d (n + \Delta n).$$

The difference

$$(AA') - (BB') = d \Delta n$$

causes a jump in the phase $\Delta\varphi$ equal to $\Delta\varphi = \frac{2\pi}{\lambda} d \Delta n.$

This discontinuity is usually small and does not exceed a few degrees. Unfortunately, the conventional optical instruments, including the microscope, are not able to detect the "dephasing" object since the variation in phase is not accompanied by a change in amplitude, and the amplitude of the rays which have passed through the impurity remains the same as for those propagating through the surrounding medium. Strictly speaking, by means of a number of devices (defocusing with a reduced aperture, indirect illumination, and so on) we can bring about a certain increase in contrast, but at the price of appreciabl^e distortion of the similarity between object and image and a decline in the resolving power of the microscope.
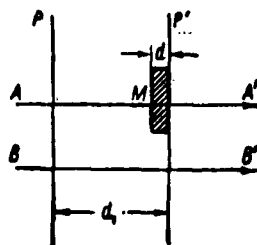

Fig. 60. Diagram showing phase contrast.

In 1931 the Dutch scientist Zernike pointed out a method of turning phase variation into amplitude variation and thereby making dephasing objects visible.

Let AB be a plane-parallel plate made of a non-absorbing material (Fig. 61). At the center O there is a transparent (non-absorbing_ impurity, the refractive index of which differs from that of the plate by the small value Δn.

Let us assume that an object is illuminated by a parallel beam of rays proceeding from the condenser CC', in the leading focus of which there is a point source of light S. Having passed through the plate AB, the rays continue as far as the microscope objective LL' without being refracted, except for those which have passed through the contour of the impurity O. The latter, on account of the discontinuity in refractive index Δn on the boundaries of the impurity, are diffracted and form a beam of diverging rays (OL, OL' etc.).

Whereas the rays which have passed through the plate AB are gathered at the focus F after refraction through the objective LL', those rays diffracted by the

edges of the impurity are gathered at the point O' after refraction, creating an image of the impurity's ᴏᴜᴛ‌ʟᴵɴᴇ, which could not normally be observed through an eyepiece against the light background created by the first beam. Thus, the observer sees a uniform field lacking all detail ~~through~~ the eyepiece.
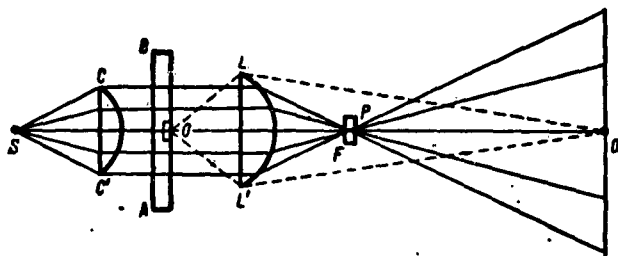

Fig. 61. Behavior of amplitude of two groups of oscillations propagated through a microscope.

The phenomenon occurring at point O on the impurity and in direct proximity to it can be interpreted in terms of the wave structure of light as follows. Let OA (Fig. 62a) be a vector mapping the luminous impulse passing through the plate. The length of this vector is a, where a is the amplitude of the oscillations. The angle with the axis (oscillation phase) is taken as zero. The light impulse produced by the impurity is mapped by the vector OB. The length of this vector is a since the amplitude is no different from that of the first oscillatory movement (neither medium absorbs), but the phase of the second differs from the first by

$$\Delta\varphi = \frac{2\pi}{\lambda} d \, \Delta n.$$

Since this angle is very small, the vector OB may be represented as the sum of the two vectors OA and OB, with AB ⊥ OA being equal to

$$AB = a \cdot \frac{2\pi}{\lambda} d \, \Delta n.$$

Consequently, the passage through the impurity, has as it were, caused the appearance of a new impulse of low amplitude with a phase differing from that of the principal impulse by $\pi/2$.
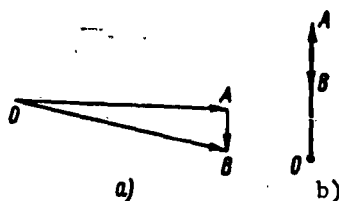

Fig. 62. Sum of amplitude vectors without phase plate.

Thus, two impulses pass through the impurity: one which is the same as the impulse passing through the plate and can be called the background; the second which is very weak and differs from the phase by $\pi/2$.

The intensity of the transmitted light is proportional to the amplitude of the resultant vector, i.e., the sum of the squares of the amplitudes of the component impulses

$$a^2 + a^2\left(\frac{2\pi}{\lambda} d\,\Delta n\right)^2 = a^2\left[1 + \frac{4\pi^2}{\lambda^2} d^2\left(\Delta n\right)^2\right]$$

and since Δn is small, the total intensity is only slightly different from a^2 (it only differs by a value of the second order of smallness). For example, if the jump in the optical path is $1/40$ of the wavelength, the variation in intensity is equal to 2.5% and practically indistinguishable.

Zernike has suggested putting a so-called <u>phase plate</u> P in the rear focal plane of the objective LL' (at the spot where the rays passing through the plate AB gather - see Fig. 61); the thickness and refractive index of this phase plate are such that when oscillations pass through it, they experience a phase change of $\pi/2$, and a reduction in amplitude through absorption of the light in it. The dimensions of the phase plate are small and it does not therefore have any effect for practical purposes on the beam diffracted by the impurity. The vector diagram in Fig. 62a now assumes the form shown in Fig. 62b. The vector OA has turned by $\pi/2$, let us say counter-clockwise, and has been reduced to dimensions close to the length of the vector AB. The latter now acts in parallel to OA, but in the opposite direction, and the resulting vector OB is equal to the difference (or sometimes the sum) of OA and OB. Let us assume that the transmission factor of the phase plate is τ; this means that the amplitude of the light oscillations passing through it is weakened by a factor of $\sqrt{\tau}$. The intensity of these oscillations as a result of the two oscillatory motions is determined by the equation

$$a^2\left(\sqrt{\tau} - \frac{2\pi}{\lambda} d\,\Delta n\right)^2 = a^2\left[\tau - 2\sqrt{\tau}\cdot\frac{2\pi}{\lambda} d\,\Delta n + \left(\frac{2\pi}{\lambda} d\,\Delta n\right)^2\right].$$

Since $\frac{2\pi}{\lambda} d\,\Delta n$ is very small compared with τ, the last summand in brackets can be disregarded and we can write down the following for the intensity of the resultant beam

$$a^2\left(\tau - 2\sqrt{\tau}\,\frac{2\pi}{\lambda} d\,\Delta n\right).$$

138

In the example given above

$$\frac{d\,\Delta n}{\lambda}=\frac{1}{40}\,.$$

Even if we assume $\tau = 1$, i.e., that the phase plate is absolutely transparent,

$$2\,\frac{2\pi}{\lambda}\,d\,\Delta n=\frac{4\pi}{40}\approx\frac{1}{3};$$

the variation in intensity attains 30% of the intensity of the background, and the impurity becomes visible. By giving τ a value causing the multiple

$$\left(\tau-2\sqrt{\tau}\,\frac{2\pi}{\lambda}\,d\,\Delta n\right),$$

to vanish, we arrive at an intensity of zero, i.e., the impurity appears dark on a bright background and the contrast between the latter and the object is maximum.

Naturally, the theory of phase contrast is not as simple as may seem the case from these highly simplified arguments. In natural fact, the light sources used to illuminate the preparations are by no means points; other phenomena occur when oscillations are propagated, which we have not considered here, and so on. The devices used to produce phase contrast are rather complicated and so far no universal method has been developed for meeting all requirements.

But the principle of phase contrast, put forward by Zernike by which the phase discontinuities are converted into amplitudes and enable us to see objects which were unobservable before, has considerably increased the possibilities of optics as regards the detection and study of an extensive class of very important objects.

As distinct from the example given above (p. 160), which is noteworthy in that the image disappears, we also find the opposite phenomenon, the appearance of an invisible object, and all that was needed was a lens a few fractions of a micron thick!

Sec. 5. Apodization

Inasmuch as the amplitude-phase filters change the distribution of illumination in the diffraction image of any object, particularly points, within very

wide limits (in theory any required distribution can be produced), we naturally try to use these filters to increase the resolving power of optical instruments.

Even before the systematic theoretical and practical study of the potentialities of amplitude-phase filters was begun, some of the properties of the simpler kind were known.

It was known that if the pvp/L in an optical system was covered over with an opaque diaphragm in such a way that only a narrow zone at the edge was left open, the resolving power would be slightly increased, but the intensity of the secondary rings was greatly increased and became almost equal to that of the central spot. This device was used on occasion by astronomers for investigating details in constellations. Here we are dealing with an amplitude filter which totally absorbs the central rays of the beam and only allows through the marginal rays.

It was also known that an optical system with a definite, though not very large spherical aberration, which plays the same part as a phase filter with the corresponding change in phase, also possessed greater resolving power than an ideal system with the same aperture, but that at the same time the intensity of the secondary rings was increased and the intensity of the central spot decreased.

Over the last decade a great deal of research has been carried out on the properties of phase, amplitude and mixed filters (principally by French scientists). The research has been made difficult by the exceptional complexity of the mathematical operations required to describe the relationship between distribution of energy in the image pattern and its distribution in the object. Although the problem has still not been definitely solved, the following conclusions can be considered as VALID.

1. Amplitude (and not phase) filters are of greatest importance for increasing resolving power.

2. By means of amplitude filters we can increase the resolving power (theoretically to any extent), i.e., we can bring about a decrease in the diameter of the central spot. However, this is always accompanied by the following two phenomena which detract from the advantage of   reducing the spot:
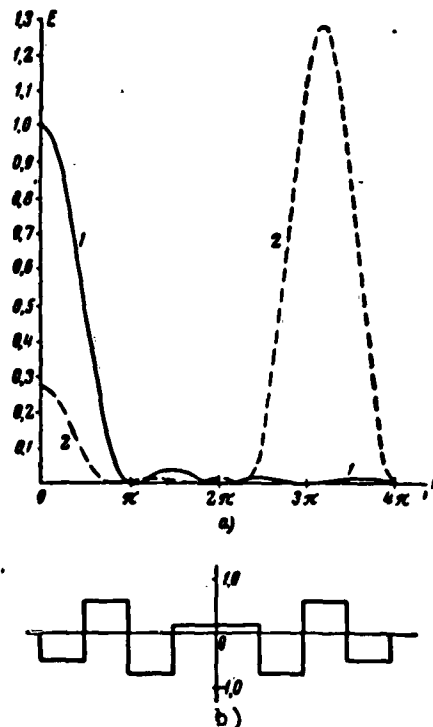
Fig. 63. Distribution of luminous energy without filter (curve 1) and with special amplitude filter (b) increasing resolving power of spectral instrument (Fig. 2).

Calculations show that these assumptions are valid. It is comparatively easy to calculate for a case in which the source of light is an infinitely thin slot directed by an optical system with an ENTRANCE PUPIL in the form of a slot of infinite height. The sides of the slot are parallel to the light source. This is the case with spectral instruments, at least with a fair degree of approximation. Fig. 64a shows the luminous energy distribution which can be obtained in the image of an infinitely thin slot if a filter with absorption obeying a certain law of variation is placed in the inlet aperture of the optical system (Fig. 64b) .

As can be seen from this figure (curve 2), the first two side maxima are totally eliminated. The following maxima can no longer be weakened, but on the contrary, are intensified, and vary considerably, too. The central band, determining the resolving power of the system is somewhat broadened and absorption

141

a) there is a considerable increase in intensity of the side rings, and the increase is more rapid than that of the resolving power. In other words, if we reduce the diameter of the central spot by a factor of 2, the intensity of the rings exceeds that of the spot by a factor of hundreds;

b) on acount of the fact that luminous energy is redistributed over the rings, the amount of luminous energy in the central spot, which in normal non-aberrating optical systems is 84% of the incident luminous energy, does not exceed fractions of a percent when we use the absorbing filter! And if the resolving power is further "increased", the amount of luminous energy incident on the central spot is disastrously decreased.

It is clear that this increase inthe resolving power is of no practical interest. $A N y w A y$, we can hardly speak of any increase in this case, since the Rayleigh criterion, which does not in effect make allowance for the effect of the rings, in view of their low intensity loses all significance. For purposes of comparison of the energy distribution curve for ordinary conditions and the distribution curve for an amplitude filter, we refer readers to Fig. 63a, which shows both curves on the same scale.

Fig. 63b shows the amplitude distribution on the filter from the left-hand to the right-hand edge. It should be pointed out that negative amplitudes are also used here; this means that the areas of the filter with our negative amplitudes are coated with thin layers producing a difference of half a wavelength between the beams passing through the main area of the filter and those transmitted by the remaining area. As can be seen from the diagram, the illumination in the center has decreased by a factor of 4, compared with the illumination obtained when there is a filter. The resolving power is increased approximately 20%; but when the argument is close to 3π, there appears a new maximum, exceeding the central one by a factor of 5!

3. Since reduction of the size of the central spot is virtually impossible, we can try to solve the opposite problem, i.e., the problem of removing the side maxima; here we should expect a certain enlargement of the central spot.

by the filter reduces the intensity by a factor of 0.6.

The spectrograph fitted with a filter, shown in Fig. 64b, has the following advantage: if there is a weak satellite close to the principal line, and the satellite is situated within the interval $n_1 n_2$, over which the side maxima have been eliminated, it will not be light filled as would happen if there were no filter. Attempts to eliminate a greater number of side maxima mean an even greater decrease in intensity of the central band and further broadening of it.

Thus, the use of amplitude filters may in certain cases be an advantage. But, generally speaking, any enforced modification (aimed at improvement) of one set of properties inevitably results in deterioration of others, usually ones which are just as important.

Dossier has shown that the use of combined phase and amplitude filters for increasing resolving power does not produce beneficial results, either.
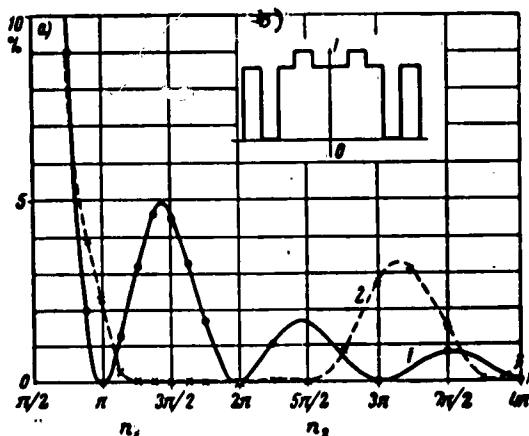

Fig. 64.

Distribution of luminous energy without filter
(curve 1) and apodizing amplitude filter (curve 2).

Sec. 6. Brightening in optics

The reflection of light from surfaces of lenses often gives rise to a great deal of trouble for people using optical instruments. Particularly inconvenienced are photographers who are thereby prevented from taking landscapes whenever the image of the sun comes within the field of vision of the objective, and even when the direction of it forms fairly considerable angles with the axis of the objective, since in the latter case the resulting photographs are to some degree fogged. It is also impossible to photograph a landscape at night if it is

143

lit by bright lights, the image of which appears in the photograph in the form of blurs due to light twice reflected from the lens surfaces. This is why it is difficult to make out anything through binoculars under these circumstances.

The reflection of light from INTER FACES with different refractive indices is described by the Fresnel laws, according to which the ratio of the reflected beam and the incident beam depends on the refractive index of the two media and the angle formed by the incident beam and the normal to the surface. In most of the optical lenses used in practice, from 4 to 5% of the incident light is reflected WHEN THE angle of incidence is not greater than 30°. Since the Fresnel law implies the impossibility of reducing the amount of reflected light below a certain level, it is not surprising that until recently it was absolutely impossible to get rid of reflected light during refraction. But the appearance of REDUCED / REFLECTION layers in the twenties, successfully developed in the USSR by Grebenshchikov and Lebedev, caused great surprise.

In actual fact, the possibility of reducing reflected light, **even without these layers**, follows from basic principles.

When light falls perpendicularly on a refracting surface, the ratio of the reflected light to the incident light, according to the Fresnel law is

$$\frac{I_0}{I} = \left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2.$$

where n_1 and n_2 are the refracted indices of the first and second media, respectively.

For a case in which $n_1 = 1$, $n_2 = 1.5$, the ratio $I_0/I = 4\%$.

Let us assume we coat the plane surface of the second medium with a layer with a refractive index $n = 1.25$ (Fig. 65a). Let us calculate the losses. In the first surface they are equal to $(0.25/2.25)^2 = 0.012$, and in the second $(0.25/2.50)^2 = 0.010$. The total losses amounted to 0.022, that is to say only 2.2% instead of 4%. In other words they have been halved. It is easy to show that if we use ten layers instead of 2, and the refractive indices are, respectively, 1.05, 1.1, 1.15, 1.20 and so on, the total losses is not greater than 0.4%, i.e., smaller by a factor of 10 than if there were no layers. Unfortunately, the system is

NOT TECHNICALLY FEASIBLE in view of the non-existence of transparent solids with refractive indices close to unity, but theoretically the idea is perfect. It is a curious fact that the thickness of the layers is of no importance; furthermore, the losses would be the same for rays with different wavelengths, since the refractive indices BARELY change with wavelength.

In actual fact the REDUCTION IN $\sqrt{}$ is quite possible by a considerably simpler method (by using one extremely thin layer), but only on account of interference.

Let AB (Fig. 65b) be a ray striking the surface MM_1 separating air ($n_1 = 1$) from glass with a refractive index n_2. Let us assume that the glass is COATED with a thin layer of transparent material with the refractive index $\underline{n}$. Some of the energy of the ray is reflected from the surface NN_1 of the layer BA_1, some of it is refracted in the direction BC and is partially reflected and partially refracted at point C. The brightness of the reflected ray $\underline{DA}_2$ is calculated from the Fresnel equation. It is easy to prove that if $n = \sqrt{n_2}$, the brightness of the ray BA_1 reflected from the layer is equal to the brightness of DA_2, reflected from the glass. If the difference in the path (BC + CD) between the rays $ABCDA_2$ and ABA_1 is half the the wavelength, then the interference of the coherent beams DA_2 and BA_1 cause the oscillations in these beams to cancel each other out, and the brightness of the reflected light is equal to zero. Hence all the energy is refracted.

The disadvantage of the "interference" method of eliminating reflected light is as follows. The total elimination of the oscillations only occurs at a definite wavelength λ_0. At other wavelengths it does not occur, and even at small deviations from λ_0 the rays are appreciably reflected. This produces a violet tinge found in COATED LENSES due to complete reflection of the blue and red rays.

It has been possible to some extent to eliminate this undesirable effect by using two or even three layers together.

Coating the lens surface with thin layers makes it possible to eliminate the

reflection of light on the refracting surfaces as well as to increase the reflective power of the surface. The $coatings$ can be used as well to increase the reflective power of metal layers - silver, aluminum or even dielectric coatings which are applied to glass surfaces when making mirrors. The use of such layers increases the reflective power of silver from 92 to 99%.
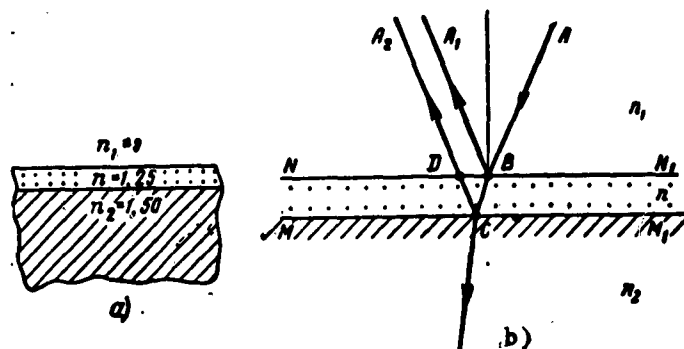

Fig. 65.   Reduction in loss during reflection: a) without use of interference effect; b) with use of this effect.

Sec. 7. Interference lenses

As Fresnel showed, the rectilinear propagation of light due to the fact that the individual surface zones of the wave created by the luminous point compensate each other, except for an extremely small center area between the source and the observer. This formulation gave Soret and Rayleigh the idea of making plates which would only allow through oscillations with an amplitude of one si and would restrain the others. Plates of this kind must possess focusing (condensing) properties .

They consist of concentric transparent and opaque rings of equal area. If h_1 is the radius of the outer circumference of the first dark ring OA_1 (Fig. 66a), it can be assumed that the "focal distance" of the zone plate is determined by the equation

$$\frac{1}{f'} = \frac{2\lambda}{h_1^2} .$$

Let 0 and 0' be two conjugate points, i.e., let them satisfy the condition

$$\frac{1}{s'} - \frac{1}{s} = \frac{1}{f'} ,$$

where s and s' are the distances between the plate and the object and image, taking

the sign rule into account (Fig. 66b). Calculations show that the optical path between the points 0 and 0', i.e., the sum of the segments OM + MO' during passage from the beginning to the end of the zone (from point 0 to A_1 or from A_1 to A_2, for example), is increased by λ , while the phase is increased by 2π . Since a phase change of the integer 2π does not alter anything in the oscillatory motion, it can be said that for point 0' the oscillations occurring on all the circumferences OA_1, OA_2, OA_3, ... etc., are in harmony, and the pulses from these circumferences are summed at the point 0'; hence 0' can be called the image of point 0. Furthermore, the image is not a good one, since the other zones of the plate, for example, those shown by a broken line in Fig. 66a, although harmonized with respect to phase, are not coordinated with the foregoing ones, and the amplitude of oscillations from these zones NO LONGER (add up arithmetically , but are summed as vectors.'
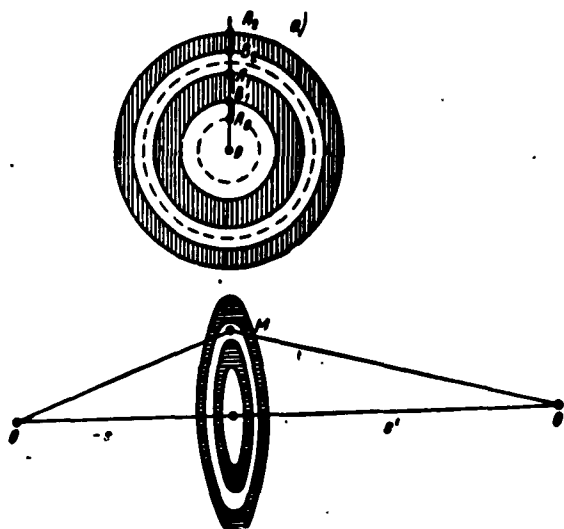

Fig. 66.   Structure of Soret zole plate

An exact calculation shows that only 10% of the luminous energy incident on the plate reaches point 0', 50% being absorbed by the opaque rings and 40% forming a number of other images. These extra images are found at all points of the axis 00' for which the optical path 00' (n is the number of the pdnt) is increased by $2\lambda_0$, $3\lambda_0$, ..., k λ_0 during passage from one ring to the next. In other words, the zone plate possesses an infinitely large set of foci determined by the equation

$$\frac{1}{f'} = \frac{2\lambda}{h_1^2} k.$$

where k is an integer.

Wood suggested the following improvement to the system. Instead of transparent layers, he suggested coatings producing a phase difference of a half wave. This stepped up the useful amount of light converging at the center of the point image 4 times, although the basic disadvantage of the zone plate, the multiplicity of foci - remained as before.

But we can go further. We can make the plate in the shape used by Fresnel for his lighthouse lenses, but on a microscopic scale; the width of the serrations AB is about one micron (Fig. 67), while the thickness of the plate varies with the height according to a law ensuring that the optical system $OM + \dfrac{MN}{n} + NO'$ for all points on the ring. When going from one ring to the next, the optical path of the ray is increased by $A \cdot jump / \lambda_0$, while the phase skips 2π which is tantamount to no jump at all.

A lens with this profile possesses the property of an ordinary lens, that is to say, it produces a sharp, single image of the point at a given wavelength and AT any position of the object. The aberration in rays of this wavelength is similar to conventional lenses, except that the spherical aberration is more easily limited and there is no curvature of the plane of the image; this is impossible in conventional lenses.

The resolving power of the lenses with a profile of this kind - which we call phase plates - is the same as in ordinary lenses with the same focal distance and diameter. The main difference between phase plates and ordinary lenses is that the former exhibit tremendous chromatic aberration, depending solely on the wavelength λ and not the material of which the lens is made. This aberration is approximately 20 times greater than in lenses made of borosilicate crown glass. Here the images in the rays, whose wavelength λ differs from λ_0, exhibit a remarkable property - their brightness is less than that of the image of the principal color. The drop in brightness, however, is small if the area of the spectrum employed (for example, the visible region) is not excessively wide. The

resolving power does not vary here.

The PHASE plate may be calculated in such a way that it compensates the residual chromatic aberration of ordinary two-lens objectives used, for example, in astronomical telescopes. This aberration, which is called the secondary spectrum, is so great that it prevents us obtaining a clear-cut picture of the celestial bodies and [is the reason why over the last few years not a single large refractor has been built, whereas an ever greater number of mirror telescopes free from chromatic aberration have been constructed and put into service. The calculation shows that the addition of a lens of a focal distance equal roughly to 7 times the focal distance of the objective, coated on one side with phase layers of a certain shape (from 30 to 50 layers according to the objective characteristics) enables us to reduce the secondary spectrum by a factor of 15 to 20. Unfortunately, the manufacture of these layers involves great technical difficulty.

We should point out once again that when studying the phase plates described here arguments based on geometric optics are a great danger; they may lead to absurd conclusions. An error resulting from such arguments increases as the width of the zones and the number of them decreases.

The use of thin coatings, described here, far from exhausts their possibilities. The possibilities resulting from the modified amplitude and phases would not have EXISTED , had not light been of an undular nature.
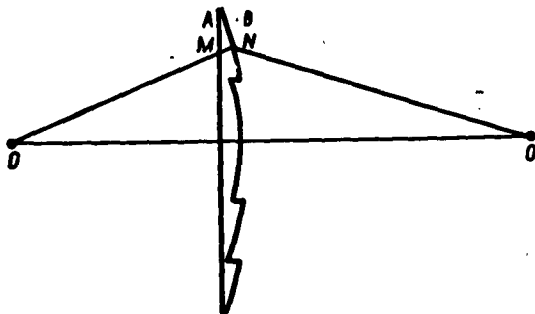

Fig. 67. Improved profile of zone plate.

Sec. 8. Some more about resolving power and the amount of information provided by an optical system

The arguments contained in Sec. 8 in Ch. IV may give rise to undue hope that it may be possible to unearth details in optical systems for which there is no

provision in the normal theory of resolving power. If we assume that the receiver possesses infinite resolving power (i.e., that it does not itself introduce any change in the pattern of the image) and that there is no interference from the atmosphere, nor aberration in the optical system, it might be thought that an experienced analyst of image patterns might be able to glean a considerably greater amount of information than   on the basis of the classical theory of resolving power.

Unfortunately, this is not the case.   Extremely COMPLEX mathematical operations based on equations resulting from the Huygens-Fresnel principle show that an image pattern, i.e., the distribution of brightness in the plane of the image, may be produced not by one object (or more exactly, not by one distribution of brightnesses in the plane of the object), but an infinitely large number of them. If the object is periodic in structure (for example, a diffraction GRATING) which is thinner than the optical system can resolve according to the Abbe elementary theory, the optical system, as it were, "refuses to allow" the structure through the objective and no matter in how much detail we study the distribution of energy in the image, its shape cannot be determined uniquely. The French physicists call optical systems of this kind "filtres passe-bas", i.e., they regard them as filters which allow through periodic structures, the frequency of which is not below a certain level.

Of lately there has been a revival of interest in a curious fact. As was stated above, there may be several (and theoretically an infinite number) of distributions in   space of the object for one and the same brightness distribution in the space of the image. But if we have additional information on the subject from other sources (let us suppose, that we know it is a double star from say a spectrographic study), the obscurity in attempting an image pattern is removed. This makes it possible to obtain further data, for example, on the distance between stars, the ratio of their brightnesses, and so on. Thus, one piece of information gives rise to another.

The problem of the "amount of information" supplied by an optical system has recently come to the attention of a number of specialists in optics. It has

been observed that the development of communications by means of which a person A
let us call him the sender - sends certain information by some means or other (Morse
code, telephone, radio, etc.), requires an answer to the following questions: a)
how much information can be transmitted in a given period; b) in what form, i.e.,
with how much distortion will the receiver B get the information. A question which
arises in this connection is, how valuable is THE information which has to be transmitted,
since IF it is of little value, for example, birthday congratulations, and arrives in dis-
torted form, no great harm is done; AFTER ALL the receipt of the news is not unexpected.
Attempts to solve these problems separately in different branches of engineering
have resulted in a new trend in a general direction which is now called the theory
of information, although it ought to be called the theory of the transmission of
information.

It has often been pointed out that the formation of images by an optical
system can also be treated as a TRANSMISSION of information. The object (for the sake
of simplicity we will consider it to be plane) represents a certain distribution of
brightness over a plane. The distribution may be regarded as a totality of inform-
ation which by means of the optical system is being transmitted to a receiver.

During the reception and transmission of luminous fluxes there are distortions.
The reason for the distortions are: energy losses in the light, aberration of the
optical system; diffraction; which turns a point into a configuration with complicated
energy distribution; the crude structure of the receiver; parasitic light entering
the receiver through secondary reflection; scatter by the framework, dust or
scratches on the surface of the lens, extraneous light reaching the surface of the
receiver, and so on. All these factors reduce the amount of information coming from
the object . For example, if the receiver is the retina of the
human eye, which has 900,000 nerve endings, the number of signals received by the
eye over a given moment of time cannot exceed 900,000, and this can only be the case
if the image of the point produced by the optical system does not exceed the size
of the individual cell through aberration or diffraction. But on the object the

number of signals is not restricted, generally speaking, and the number of them
which the object can dispatch cannot easily be calculated.

As a simplification it can be considered that the object represents a
lattice or grid (two lattices at right angles). In a more general case we can
represent the brightness B (xy) of a point on a flat object in the form of a
Fourier series with respect to two variables x and y, or else by other methods.
If it is known how any point on the object is reproduced in the plane of the image
(which is only feasible in very simple cases, in particular when the optical
system is ideal), it is possible at least theoretically to calculate the distribution
of illuminations in the plane of the image. In practice, this problem is solved
by very laborious and cumbersome methods.

The distribution pattern of illumination in the plane of the image (Ronchi
suggested calling it "aetherial" since it cannot be observed in the pure form under
any circumstances) is by itself a distorted reproduction of the object of a certain
scale, but it is just the unreal or "aetherial" nature of it which hampers calculation
of the number of independent signals describing it; furthermore, the number of
signals is of no practical interest as long as the receiver receiving the image is
unknown.

Knowing the properties of the receiver, its structure, the sensitivity of
individual parts of it, the threshhold of sensitivity to contrast of neighboring
parts or groups, we can determine the number of separate "signals" which are able
to be received and transmitted by the receiver. The ratio of these signals to the
signals dispatched by the object may serve as a measure for the distortion introduced
by the system "optical instrument - receiver".

However, the formation of any image cannot be covered entirely by any means
by the theory of information, since, apart from the number of measurements and
dependents on time, there are fundamental differences between this formation and
the transmission, say, of a conversation by telephone.

The theory of information makes it possible to consider optical images from
ew angles which are of definite interest, and because of the analogy between

different vibratory processes, enable us to transpose the results of one branch of engineering to another. It is true that not all concepts pertaining to the theory of information and associated with the trans ^(mission) of information can be mechanically transposed to the theory of optical images. For example, the function introduced by Shannon under the name of entropy and similar in appearance to the entropy which in thermodynamics indicates the direction in which a conversion or transfer of energy should move, meets with difficulties if we try to transpose it to optics where the aim of the function is to give a numerical value to the concept of "information value".

Any concept, of course, can be introduced into optics with a certain degree of conventionality, but the benefit to be derived from this artificial transfer is questionable. It should be pointed out here that the theory of information has not revealed or cannot reveal any new possibilities either in improving the quality of images or in man's ability to derive greater advantage from the analysis of the image pattern. Thus, the theory cannot extent the frontiers of the "possible" in optics, but merely puts a new language at the disposal of the investigator. Nevertheless, cases are known in which a new language leads to new ideas. The language of the theory of information has enabled the Italian optics specialist, Toraldo, to derive a new criterion of quality for photographic objects.

CONCLUSION

Despite the sketchiness of the material contained in this book, it can nevertheless be used to draw general conclusions on the limitation of optical systems. These limitations can be ascribed to two spheres: 1) macroscopic phenomena, i.e., phenomena produced by light sources with large linear or angular dimensions, and 2) microscopic phenomena in which the light sources are very fine structures studied under a microscope, or heavenly bodies, the angular dimensions of which are not greater than fractions of a second.

In the first sphere the principal limitation on the structure of light rays is the law of conservation of energy, or in the language of the optical experts, the Lagrange-Helmholz law. This law stands in the way of realization of dreams of long-range combustion until light sources thousands of times brighter than the sun are developed, and removes all hope of constructing concentrating devices, BEAMING strong LIGHT through narrow openings OR even calculating ultra-rapid photographic lenses, which is also a favorite subject for inexperienced inventors. Furthermore, everything that has been said in connection with long range combustion by means of heat or visible rays is true of any rays whether infra-red or ultra-violet, and generally speaking, OF any type of radiant energy.

The irreversibility principle categorically refutes the transformation of scattered into beamed light, which is so alluring for television designers; it excludes the possibility of developing lighthouse optical systems similar to the one described in detail in Ch. III.

The vibratory nature of luminous phenomena is beginning to MAKE AN APPEARANCE ⟨in the microworld and the concept of wavelengths is beginning to make itself felt. The comparatively short wavelengths lying far beyond the threshhold of resolution of the human eye HAVE fortunately not impeded the development of the simple and neat theory of geometric optics, which has enabled us to construct all the wonderful optical instruments which science has used to take such giant strides forward in unraveling the secrets of nature. But these instruments have enabled us to detect their enemies - diffraction and interference - which are results of the vibratory nature of

the propagation of light. It seemed that a point could never be projected as a point, that all optical instruments produce distorted images of objects being viewed and that we would have to learn to recognize the objects in much the same way as a reconnaisance scout picks out carefully camouflaged objectives. Naturally, a science of this kind is not an easy thing, and many an investigator, astronomer or biolgist has been led astray by taking too simple an approach to the picture he was studying.

The attention of the readers has also been drawn to the fact that the concept of resolving power is conditional and that different things can affect it to different degrees. Certain methods based on the undular nature of light enable us to increase this power; but no great hopes should be placed on these methods. Although in principle, it is possible to produce an image of a point by an optical system in different forms, using amplitude filters with a particular density distribution, in order to do so we have to sacrifice practically all the incident luminous energy and leave a mere fraction of it. That is why this approach cannot be considered in any way promising and it would be unwise for inventors to direct their efforts towards it. Furthermore, the great mathematical difficulties guarding the path can themselves be considered reliable defense against casual attempts to abuse these possibilities.

A number of limitations, this time of a technical nature rather than a question of principle, are by aberrations of optical systems. Because it is impossible to eliminate them by simple methods (by using a minimum number of lenses, mirrors or other parts) we cannot build optical systems which have a large relative aperture and a large angle of vision at the same time, maintaining a high-grade image over the whole field. That is why we cannot construct short optical systems producing large-size images and satisfying the several conflicting requirements at the same time, for instance, problems of size. This sphere of computation is also of interest to inventors wishing to overcome obstacles which have so far been insurmountable. But it is usually only the experienced specialists who engage in such research, and iF quite a number of pages are devoted to this matter, it

155

is not in order to DISCOURAGE young inventors from ~~temper~~ WASTING THEIR TIME, BUT RATHER TO temper the apetites

of over-demanding clients reluctant to face unwritten, though actual and inevitable facts.

We have talked separately, and admittedly rather briefly, with certain "optical" systems and instruments using x-rays and electron beams. We have left aside "acoustic" optics and radio optics, since on account of the great wavelength, the undular nature of radiation predominates over geometric radiation, and the Huygens-Fresnel principle translated into the language of new vibrations is no longer a secondary rider to the laws of refraction and reflection, but, on the contrary, is becoming a basic one. This is bringing about great changes in the design of optical systems transmitting and receiving acoustic and radio signals.

* *
*

The author hopes that this book will help inventors to channel their efforts in the right direction by warning them of the many dangers lying in wait along the difficult path of invention, and that it will help them to save time and energy which are sometimes so easily wasted on attempts to circumvent the basic laws of optics. The book will also help them to recogize these laws in the complicated instances in which they are happily disguised by second-rate facts which at first sight seem important. Attention has also been given in the book to the latest views on certain aspects of optics, developed over the last few years and not yet fully established. A study of them will make it possible to understand the basic principles of optics more thoroughly and to study the older problems from a new standpoint.

APPENDIX

FIRST PROOF OF LAGRANGE-HELMHOLZ LAW

Because of the great importance of the Lagrange-Helmholz law, we will give
the proof of it here. We will first show that if losses of light during passage
through an optical system are disregarded, the brightness of the beam divided by the
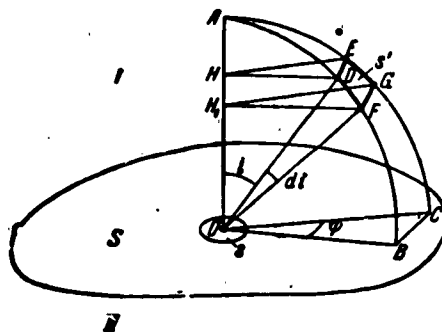square of the refractive index is constant along the whole beam.

Fig. 68. Generalization of law of refraction.

Let us consider the element s of the focus S (Fig. 68) separating two
media I and II with refractive indices n and n'. Let us assume that a beam P
limited by the elements s and s' impinges upon the surface S. A sphere with a
radius of unity is drawn from the point O of the element s. Let OA be normal to the
element s and let DEGF be an element cut from the surface of the sphere by the beam
P. The boundaries DE and FG of this beam are determined in the following way. Let
us draw two planes OAB and OAC through OA; these planes form a very small angle ϕ.
Let us consider two planes parallel to each other and to the element s, and
perpendicular to the normal OA and intersecting the latter at points H and H .

Let us calculate the flux passing through the light ray bounded by the
elements s and s'. To do this let us first determine the area of the element DEGF,
which on account of very small size can be regarded as a rectangle; its area is
therefore equal to DE ⋅ EG. Let us use i to designate the angle formed by the rays
OD and OE and the normal, and i + di to designate the angle formed by the rays

OF and OG. Since the radius of the sphere is unity, we may assume that EG = DF = di. The side DE as part of the circumference with a radius HD, is equal to HD $\cdot \psi$, but

$$HD = OD \cdot \sin i = \sin i,$$
$$DE = \psi \sin i.$$

The side FD = OD $\cdot$ di = di. The area DEGF is equal to ψ di sin i. The element DEGF is normal to the rays of the beam P, and its area divided by the square of the radius, i.e., by unity, is equal to the solid angle ω determining the angle of radiation of the element s. The equations in Sec. 3 of Ch. I easily give us the following expression for the flux

$$\Phi = Bs \cos \theta \cdot \omega,$$

where θ is the angle between the normal to the element s and the axis of the beam, and is equal to i. Then

$$\Phi = Bs \cos i \cdot \omega = Bs \cdot \cos i \sin i \, di \cdot \psi.$$

After refraction on the element s of the interface S, the beam intersects the sphere with radius equal to unity a second time. Let us use B' to designate the brightness of the refracted beam. The flux of this beam can be calculated exactly in the same way as for the incident ray, to wit,

$$\Phi' = B's \cos i' \sin i' \cdot \psi' \, di',$$

where B', i', di' and ψ' designate the quantities corresponding to B, i, di and in medium I. The quantity ψ' is equal to ψ, since, according to the first law of Descartes, the incident ray when refracted and normal to the element s of the refracting surface are in one plane; the rays OD and OF lie in the plane AB and remain in it after refraction; the rays OE and OG are in the plane AC before and after refraction.

Now let us apply the law of refraction of light
$$n \sin i = n' \sin i'.$$

Let us square it
$$n^2 \sin^2 i = n'^2 \sin^2 i'.$$

Differentiating the relationship, we get
$$2n^2 \cos i \sin i \, di = 2n'^2 \cos i' \sin i' \, di'. \tag{18}$$

158

We should point out that if we disregard losses, the refractive stream is equal to the incident stream, and therefore $\bar{\Phi} = \bar{\Phi}$, or dividing both streams by the common quantities - the area s and the angle ϕ - we get

$$B \cos i \sin i\, di = B' \cos i' \sin i'.di'.$$

Dividing by Eq. (18) we get

$$\frac{B}{n^2} = \frac{B'}{n'^2}.$$

(19)

This law can be applied to any refracting surface; hence it can be said that all the way along the beam of light the ratio of the brightness B to the square of the refractive index is constant and equal to the brightness B of the beam in air.

If the variation in the refractive index occurs smoothly and not by jumps, the proven relationship is still valid, since a continuous variation is tantamount to a large number of very small jumps in the index.

Now let us assume that a square area O (Fig. 69), the length of the sides of which is l, is projected by an optical system in the form of the area l'. Let us also assume that the area obeys the Lambert law, that the entrance pupil pp is circular with its center at the axis, and that its radius can be seen from the area O at an angle u. The solid angle ω which restricts the beam radiated by the area O is equal to $\pi \sin^2 u$. The flux Φ passing through the optical system is

$$\Phi = \pi B' l^2 \sin^2 u,$$

where B is the brightness of the area O. After refraction through the system the flux may be written down in the form

$$\Phi = \pi B' l'^2 \sin^2 u',$$

so that

$$B l^2 \sin^2 u = B' l'^2 \sin^2 u'.$$

(20)

Dividing Eq. (20) by (19) we get

$$n^2 l^2 \sin^2 u = n'^2 l'^2 \sin^2 u'$$

or

$$n l \sin u = n' l' \sin u'.$$
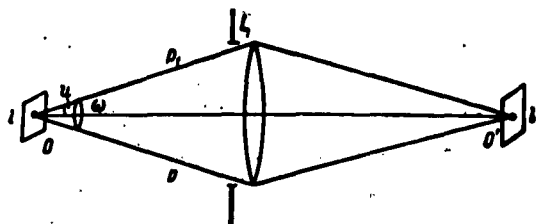
(21)

This is indeed the Lagrange-Helmholtz law.

Fig. 69. Proof of the Lagrange-Helmholz law.

On account of the fundamental importance of the Lagrange-Helmholz law, we will prove it in another way on the basis of completely different (undular) principles.

Second Proof of Lagrange-Helmholz Law

This proof follows from the well-known Fermat principle. The optical path between two points A and A' (Fig. 70) located in different media with refractive indices n, n , and n_3, respectively, is determined by the equation

$$(AA') = n_1 AB + n_2 BC + n_3 CA.$$

.e Fermat principle implies that the optical paths for the two rays ABCA' and AB C A' emerging from the point A and converging at point A' are equal. This follows, incidently , from the Huygens wave theory as well.
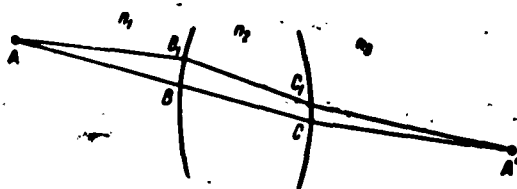

Fig. 70. Formulation of Fermat principle.

Let us now go on to prove the Lagrange-Helmholz law. Let O be a point, he image O' of which is stigmatic, i.e., all the rays proceding from point O intersect at point O' (Fig. 71). Let us see under which conditions point O infinitely close to point O in the direction perpendicular to the axis is represented stigmatically at point O'. Let us draw the ray AOB through O so that it forms angle

u with the axis. After refraction the ray proceeds in the direction B'O'A' and forms an angle u' with the axis. Let us draw the ray AO_1C through point O_1 so that it forms A_N angle u + du with the axis, hardly different from u. This ray after refraction moves in the direction C'O'A', forming with the axis the angle u' + du', hardly different from u'. According to the Fermat theorem, the optical paths are
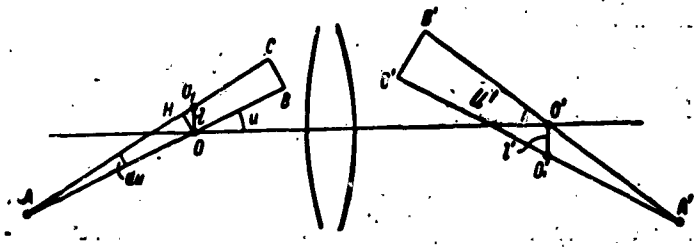

Fig. 71 Proof of the Lagrange-Helmholtz law.

equal to each other, from which we get

or
$$(ACC'A') = (ABB'A'),$$

$$nAO_1 + (O_1O_1') + n'O_1'A' = nAO + (OO') + n'O'A',$$

from which $n(AO_1 - AO) + [(O_1O_1') - (OO')] + n'(O_1'A' - O'A') = 0.$

$$(22)$$

Taking point A as the center, let us describe a circumference with radius AO intersecting the ray AC at point H. On account of the smallness of the angle du, we can replace the arc of circumference by the line OH. We then get from the triangle OHO

$$HO_1 = OO_1 \sin u = l \sin u,$$

or

$$AO_1 - AO = HO_1 = l \sin u.$$

Similarly, we get for 0' and 0'

$$A'O' - A'O_1' = l' \sin u'.$$

Substituting into Eq.(22) we get

$$n'l' \sin u' - nl \sin u = (O_1O_1') - (OO').$$

Since O' is the image of O, and O'_1 is the image of O_1, the optical paths (OO') and $(O_1O'_1)$ need not be functions of u. In the partial case, in which u = 0, the angle u' = 0 and the left-hand side of the equation is equal to zero, while the right-hand side, which is constant, should also be equal to zero. This gives us the Lagrange-Helmholtz law

$$n'l' \sin u' = nl \sin u.$$

DISTRIBUTION LIST

DEPARTMENT OF DEFENSE	Nr. Copies	MAJOR AIR COMMANDS	Nr. Copies
		AFSC	
		SCFTR	1
		ARO	1
HEADQUARTERS USAF		ASTIA	10
		TD-Bla	3
AFCIN-3D2	1	TD-Blb	3
		AEDC (AEY)	1
		SSD (SSF)	2
		ESD (ESY)	1
OTHER AGENCIES		RADC (RAY)	1
		AFSWC (SWF)	1
CIA	1	AFMTC (MTW)	1
NSA	2		
AID	2		
OTS	2		
AEC	2		
PWS	1		
PGE (Requester) (STANG)	2		
RAND	1		